THE UNIVERSITY *of* EDINBURGH

Dscam Gene Expression in Invertebrate Immunity: alternative splicing in response to diverse pathogens

Paul H. Smith

PhD Thesis

University of Edinburgh

2012

Declaration

I declare that this thesis has been composed by myself and is entirely my own work, except for the collaborative input cleared outlined at the beginning of each chapter. The work has not been submitted for any other degree or professional qualification.

Paul H. Smith

The following paper has arisen from work described in this thesis:

Smith P, Mwangi J, Afrane Y, Yan G, Obbard D, Ranford-Cartwright L, Little T: **Alternative splicing of the *Anopheles gambiae* Dscam gene in diverse *Plasmodium falciparum* infections**. Malaria Journal 2011, **10**(1):156.

Thesis Abstract

Invertebrates show enhanced immunity and even specific primed immunity in response to repeat infections, analogous to vertebrate adaptive immunity. Little is known of the mechanism for this phenomenon, or which molecules are involved. A candidate gene for the underlying mechanism for a pathogen-specific response in invertebrate immunity is Down syndrome cell adhesion molecule (Dscam). Dscam can produce thousands of different protein isoforms through the mutually exclusive splicing of many exon variants contained within variable regions of the gene. It is an important receptor of the invertebrate nervous system but has been implicated in having a role in immunity. Dscam has been shown to be involved in phagocytosis across members of the Pancrustacea, and it has been reported to respond in a pathogen-specific manner in mosquitoes and crayfish. In this thesis, I have investigated the splicing of Dscam in response to diverse pathogens in different host species.

In the *Anopheles* mosquito, I cloned and sequenced a fragment of Dscam spanning across two of its variable exon regions to enable me to detect mutually exclusively splice variants and their associations in different treatments (Chapter 2). I discovered that the expression diversity of the hypervariable Dscam is higher in parasite-exposed mosquitoes. In Chapter 3, I extended the study to the more experimentally amenable *Drosophila* fruit fly. A new Illumina-based sequencing assay was developed and implemented to examine more closely Dscam expression in response to diverse pathogens. The new method successfully quantified non-random expression of Dscam variable exons 4 and 6. I also describe a small but detectable effect of pathogen-exposure on the expression of Dscam exon 4 variants. In Chapter 4, I expanded the work of Chapter 3 to study tissue-specific Dscam expression in response to well-characterised immune elicitors of *Drosophila*. I describe how exon 4 variants were expressed in a tissue-specific manner, but not exon 6 variants. I also found a small exon 4-by-tissue-by-pathogen effect, which although detectable, did not dominate over

the tissue effects. Finally, in Chapter 5, I turned to the crustacean, *Daphnia*, to study Dscam expression in a natural host-parasite interaction and in a clonal organism. I describe the non-random expression of exons 4 and 6, and another small effect of pathogen-exposure on the expression of Dscam exon 4.

My work aimed to further investigate the putative pathogen-specific alternative splicing of the hypervariable Dscam receptor. The data presented quantified the constitutive expression of Dscam exons 4 and 6 in different pancrustacean species. The data also suggest that infection-responsive splicing of Dscam may occur but that effects are small, and may be diluted within the background of the highly important Dscam expression of the nervous system if they exist at all. The study supports the high-throughput sequencing method for future studies of alternative splicing and Dscam expression.

Acknowledgements

"If you want to improve, be content to be thought foolish and stupid..." ***Epictetus***

Firstly, eternal gratitude goes to Prof. Tom Little and Dr Darren Obbard. Their belief, support, and encouragement got me through. I would not have accomplished this without them. Thank you!

I am very thankful for my friends and colleagues at Edinburgh. Thank you to Seanna, Pedro, Phil, Carolyn, Claire, Ben, Desiree, Rebekah and Marisa. You're all very special people and I wish all of you the very best for the future.

To my son Sam, you blessed me with your dramatic arrival smack in the middle of my research, and I hope you look upon this as an example of what hard work and determination can achieve even during challenging times. To my dear mother, the one constant in my life, you always believe in me and in everything I do, thank you.

Finally, I'm very grateful to BBSRC for their financial support.

In memory of my Dad and Grandad, two courageous men lost during my PhD…I like to think I'd have done them proud.

"A scientific man ought to have no wishes, no affections, - a mere heart of stone."

Darwin

Contents

1. General introduction

I wrote this chapter with comments on drafts from Darren Obbard and Tom Little.

1.1. Overview of host-parasite interaction and immunity

Parasitic organisms, defined broadly to include pathogenic bacteria, viruses, fungi, protozoa and helminths, can negatively affect their hosts' fitness. Often causing death or sterility, parasites can exert a selective pressure on populations (Combes 2000; Schlenke and Begun 2003). The Metazoa have evolved an immune system to detect and destroy parasites (Du Pasquier 2006), thus increasing the chances of a host organism resisting debilitating disease and successfully reproducing. Conflict between hosts and parasites has led to the evolution of a complex frontline immune system for both invertebrates and vertebrates, but also to an anticipatory immune system for the jawed vertebrates (Du Pasquier 1992; Cooper and Alder 2006).

The two subdivisions of the metazoan immune system are the *innate* response, shared by invertebrates and vertebrates, and the *adaptive* response, found only in some vertebrates. The adaptive immune system is the best studied subdivision of immunity (Hoffmann, Kafatos et al. 1999). It is exclusive to the gnathostomes (the jawed vertebrates), and exhibits a highly-specific pathogen-recognition capability. The highly-specific aspect of adaptive immunity is attributed largely to specialised cells and the clonally selected functional antibodies they produce (Du Pasquier 2001). Antibodies can effectively identify and neutralise infectious organisms previously exposed to the immune system. Thus, antibodies bestow some animals with an anticipatory response to subsequent exposure of repeat infections. The innate immune system on the other hand is common to both invertebrates and vertebrates, and similarities between invertebrate and vertebrate recognition and effector mechanisms suggest a common ancestry of the immune defences (Fig.1) (Hoffmann, Kafatos et al. 1999; Cooper and Alder 2006).

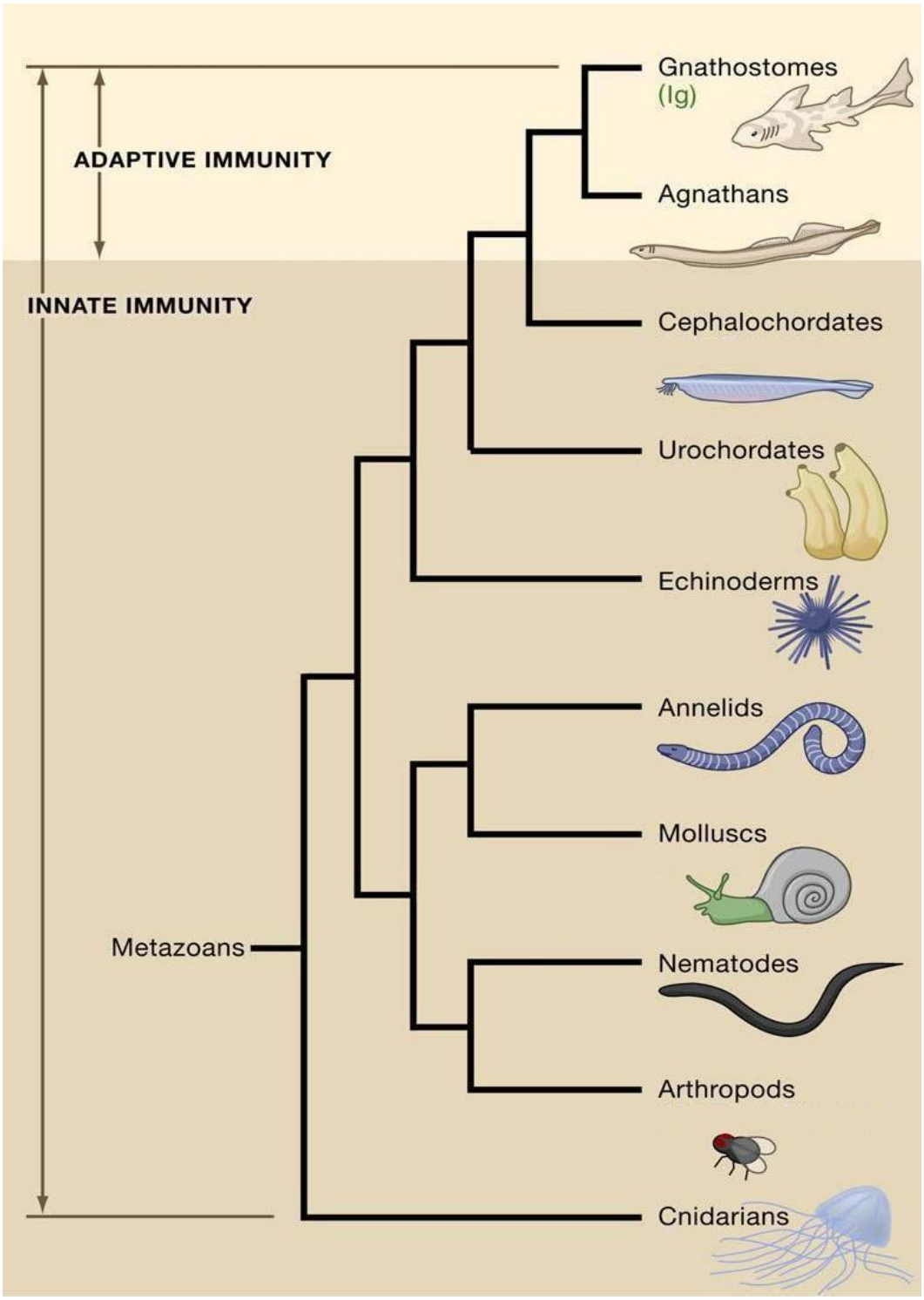


Figure 1: Phylogenetic tree indicating the theoretical emergence of vertebrate adaptive immunity. Innate immunity is common to both invertebrates and vertebrates, whereas adaptive immunity conferred by antibodies (Ig) is exclusive to the gnathostomes only. Image adapted from (Cooper and Alder 2006).

Innate immunity is a rapidly acting frontline response and probably eliminates the majority of infectious organisms and in a non-specific manner. It consists of constitutive as well as inducible defence mechanisms. Constitutive mechanisms include anatomical barriers such as epithelia and some antimicrobial peptides like the vertebrate antibacterial enzyme, lysozyme. Inducible defences are only activated in response to injury or the presence of infectious organisms. They can include mediators of inflammation, proteolytic complement proteins, non-specific reactive intermediates, components of the melanization process, and phagocytosis (Hoffmann 1995; Medzhitov and Janeway 2000). Invertebrates, having no antibodies, appear to lack the type of adaptive immunity seen in vertebrates (Kurtz 2004). Thus, invertebrates seem to rely solely on the intrinsic components of innate immunity to fight infection.

1.2. Mechanisms of invertebrate immunity

It is thought that recognition of pathogens in invertebrate immunity happens with a limited repertoire of germ-line encoded pattern recognition receptors (PRRs) (Janeway and Medzhitov 2002). PRRs detect and respond to pathogen-associated molecular patterns (PAMPs): conserved molecular motifs of many types of microbes (Medzhitov and Janeway 2000). These receptors can be cell-bound or secreted, and principally function to trigger both cellular and humoral responses to infectious organisms (Medzhitov and Janeway 2000; Beutler 2004).

Cell-mediated immunity in invertebrates is conferred by haemocytes, the equivalent to white blood cells in vertebrates. An effective cellular response mediated by haemocytes against parasites is phagocytosis (Stuart and Ezekowitz 2008). Phagocytosis is the process of engulfing and degrading microorganisms. When a cell-bound PRR binds with PAMP, the

interaction triggers the modification of the intracellular structure of a phagocyte and internalisation and destruction of the target follows (Lemaitre and Hoffmann 2007). A prerequisite of phagocytosis in invertebrates may be the opsonisation of pathogens by immune factors (Ulvila, Vanha-Aho et al. 2011). Opsonisation is a process of labelling pathogenic surfaces with immune proteins for phagocytic recognition. In mammals, opsonisation is carried out by complement proteins, which are an array of proteins which can interact to create a membrane attack complex for cell-killing, but can also label pathogens before binding cell-surface receptors to mediate phagocytosis (Ulvila, Vanha-Aho et al. 2011). For example, insect thioester-containing proteins (TEPs), which share homology with mammalian complement proteins, have been shown to be strongly up-regulated in response to infection in *Drosophila* and *Anopheles* (Blandin, Shiao et al. 2004), and required for the efficient uptake of bacteria in *Drosophila* cell culture (Stroschein-Stevenson, Foley et al. 2006). As such, phagocytosis is an important immune response in invertebrate immunity, and may be partly facilitated by other immune recognition molecules.

Significantly, PRRs play a leading role in the activation of complex signal transduction pathways leading to humoral immunity (Medzhitov and Janeway 2000; Christophides, Vlachou et al. 2004). Important signalling pathways of the innate immune system in arthropods for example, include the Toll, Imd and Jak/STAT pathways, the activation of which leads to the production of effector genes specific to the individual pathway (Fig.2).

The Toll pathway is primarily triggered by constituents of gram-positive bacteria and fungi (e.g. bacterial peptidoglycan, yeast zymosan) (Hoffmann 2003; Akira and Takeda 2004), and through the transcription factor Dif, up-regulates the production of antimicrobial peptides (AMPs) such as Drosomycin (Fig.2) (Lemaitre and Hoffmann 2007). The Imd pathway responds principally to elements of gram-negative bacteria (e.g. lipopolysaccharide) (Tzou,

De Gregorio et al. 2002) and triggers the production of the AMP Diptericin via the transcription

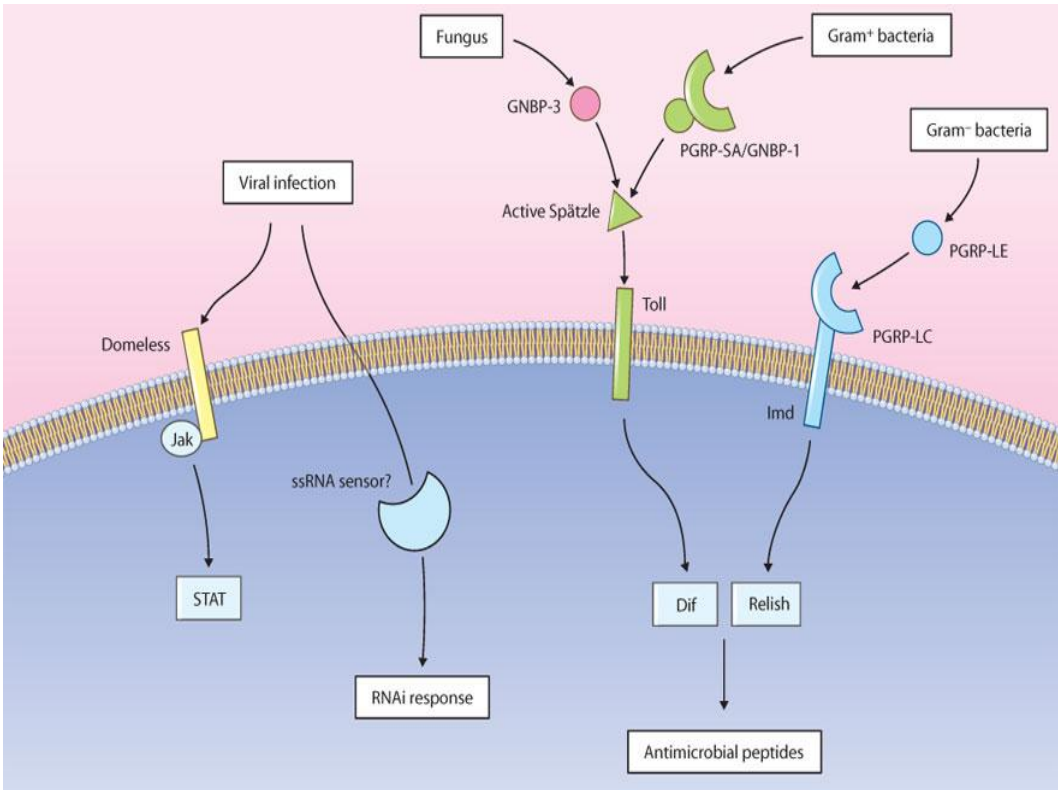


Figure 2: Signalling pathways of the innate immune response in arthropods. Three important pathways of the innate immune response in arthropods are Toll, Imd and Jak/STAT, each responding to different pathogen types. The Toll pathway primarily responds to constituents of fungi and gram-positive bacteria, and downstream signalling of Toll results in the production of antimicrobial peptides such as Drosomycin via the transactivator Dif. The Imd pathway principally responds to elements of gram-negative bacteria, and downstream signalling of Imd results in the production of antimicrobial peptides such as Diptericin via the transcription factor Relish. Finally, the Jak/STAT pathway produces anti-viral factors in response to the presence of viral proteins involving cytokine receptors like *Drosophila* Domeless. Image taken from (Fritz, Girardin et al. 2006).

factor Relish (Fig 2) (Lemaitre and Hoffmann 2007). The Jak/STAT pathway, although less well characterised than the Toll and Imd pathways, appears to respond to the presence of viral proteins by triggering the production of anti-viral factors (Fig.2) (Dostert, Jouanguy et

al. 2005; Cherry and Silverman 2006). Therefore, not all invading pathogens are treated in the same way by the invertebrate immune system, and as different types of pathogens trigger different responses, some level of specificity exists within the invertebrate innate immune response.

1.3. Evidence for greater specificity in invertebrate immunity

How specific is the invertebrate immune response, and what molecules and cells confer specificity in invertebrates? The concept of vertebrate-like specificity in invertebrate immunity was initially rejected after the failure to detect analogous components in the more 'primitive' immune system (Klein 1989) (but see Cooper, Rinkevich et al. 1992). However, the scarcity of similar components of a vertebrate-like anticipatory response cannot rule out a functional equivalent in invertebrates (Little, Hultmark et al. 2005). Despite the availability of a limited repertoire of PRRs, there is much evidence for a high degree of molecular diversity and specificity in invertebrate immunity (discussed in detail below).

1.3.1. The immune system can be 'primed' for a generalist or a specific response

Immune responses can be long or short-term, general or specific, and even within and between generations. For example, immune responses can be general (i.e. non-specific) and short-term, e.g. a brief up-regulation of the highly-reactive non-specific antimicrobial compound nitric oxide (James 1995), or longer-term, e.g. the persistence of the antimicrobial peptide defensin (Bulet, Cociancich et al. 1992). Additionally, a long-term specific immune response within a generation can occur, such as the life-long immunity in humans facilitated by memory lymphocytes and antibodies. Moreover, transgenerational priming of the immune system can also occur. For example, mammals can pass maternal antigen-specific antibodies to offspring via colostrum conferring short-term specific immunity between generations.

Thus, it is important to understand clearly the difference between general and specific immunity and how these types of responses can be conferred, especially in invertebrate immune systems which are devoid of long-lived lymphocytes and highly-specific antibodies.

1.3.2. Invertebrates show some evidence of 'enhanced immunity' following immune challenge

The first hint of a more complex invertebrate immune response was provided by Boman et al (1972), who found evidence of *enhanced immunity* in *Drosophila*. The researchers discovered that after an initial immune challenge, a long-lasting up-regulation of antimicrobial peptides in the insect's haemolymph (insect circulatory fluid; equivalent of vertebrate blood serum) conferred protection against subsequent challenges (Boman, Nilsson et al. 1972). Since this discovery, more examples of similarly-defined enhanced immunity have been presented (Hildemann, Raison et al. 1977; Rheins, Karp et al. 1978; Cooper and Roch 1986; Hartman and Karp 1989; Moret and Siva-Jothy 2003; Cong, Song et al. 2008) (also see Sequeira, Tavares et al. 1996). These observations could not rule out a sustained up-regulation of general immune mechanisms, which is quite different from parasite-specific memory following initial infection. However, indications of a more specific innate immune response have been also been reported (discussed below).

1.3.3. Invertebrates show evidence of non-specific and specific 'transgenerational priming' upon infection

While the adaptive value of maternal transfer of specific immunity in mammals and birds has been well acknowledged (Grindstaff, Brodie et al. 2003; Baintner 2007; Boulinier and Staszewski 2008), the maternal transfer of immunity to offspring in invertebrates has been controversial. Moret et al (2001) discovered that offspring from bacterial lipopolysaccharide (LPS)-primed bumblebees had increased constitutive levels of phenoloxidase; an important

component in the melanization process in insects and crustaceans (Soderhall and Cerenius 1998; Liu, Jiravanichpaisal et al. 2007). However, this enhanced prophylaxis effect across generations did not strictly demonstrate pathogen-borne specificity, and the possibility of socially-conferred immunity could not be ruled out. Indeed, horizontally transferred parasite-resistance is a phenomenon found in other animals including beetles and termites (Barnes and Siva-Jothy 2000; Traniello, Rosengaus et al. 2002). In termites for example, individuals had increased resistance to infection when placed in direct contact with previously immunized nest-mates (Traniello, Rosengaus et al. 2002).

In the flour moth *Ephestia kuehniella*, Schmidt and co-workers discovered that experimentally-induced tolerance of a bacterial endotoxin could be vertically transferred to offspring. After a few generations, insects were tolerant of endotoxin levels which would have otherwise been lethal to the starting colony (Rahman, Roberts et al. 2004). The tolerance effect correlated with significantly increased melanization reactions, and hence with elevated levels of immunity. Thus, it seems enhanced immunity can be vertically transferred to invertebrate offspring without the provision of specificity.

Finally, in the crustacean *Daphnia magna*, evidence for the maternal transfer of parasite-specificity was found following the artificial infection of adult *Daphnia* with two different strains of the endoparasitic bacterium *Pasteuria ramosa* (Little, O'Connor et al. 2003). *Daphnia* offspring were initially exposed to one *Pasteuria* strain or the other, and following a second homologous immune challenge (i.e. with the same pathogen), but not with a heterologous immune challenge, *Daphnia* newborns had a relatively enhanced fitness. Thus, this is strongly suggestive of strain-specific transgenerational priming conferred from mother to offspring.

1.3.4. Invertebrates may have 'specific memory'

A strain-specific response can be transferred to offspring after 'priming' the mother with a specific immune challenge (Little, O'Connor et al. 2003), but what about a sustained and specific memory response to pathogens during the lifetime of an individual invertebrate? Studies have provided evidence showing that, following a second homologous immune challenge during a single lifetime, some invertebrates show either a significant reduction in the rate of re-infection (e.g. in the copepod (Kurtz and Franz 2003), and the scallop (Cong, Song et al. 2008)), or a significant increase in survival rate (e.g. in the cockroach (Faulhaber and Karp 1992), the red flour beetle (Roth, Sadd et al. 2009), and the shrimp (Witteveldt, Cifuentes et al. 2004)), when compared to the administering of a subsequent heterologous challenge. In one study, bumblebees were exposed to either homologous or heterologous immune challenges with different types of bacteria. Those receiving homologous re-exposures of bacteria showed significantly higher survival rates than those given heterologous re-exposures, even several weeks after the initial challenge (Sadd and Schmid-Hempel 2006). In addition, following zone-of-inhibition assays, the authors confirmed that this specific response was independent of any protection from lingering antimicrobial peptides. Thus, studies based mainly on the assay of non-immune parameters (i.e. survival, rate of re-infection) have been strongly suggestive of specific memory in invertebrates.

A recent study in *Drosophila* may provide the best evidence for within-generation invertebrate immune specificity and memory to date. Flies were injected with sub-lethal doses ('priming doses') of the gram-positive bacterium *Streptococcus pneumoniae*, and the natural fly fungal pathogen, *Beauveria bassiana*, followed by a second and otherwise-lethal dose one week later. The results showed that the priming dose had a protective effect. Flies which were primed prior to being given a lethal dose of *S. pneumoniae* or *B. bassiana* either died more slowly or at the same rate as wounded controls. The authors concluded that the initial smaller doses provided a protective effect by priming the fly's immune response to the

specific pathogen, an effect which also persisted for the life of the insect. Additionally, *S. pneumoniae*-primed flies were tested for their ability to kill the gram-negative bacterium *Escherichia coli*. They found no difference between naïve PBS-injected flies and *S. pneumoniae*-primed flies in their ability to clear *E. coli*. This confirmed the remarkable observation that priming with *S. pneumoniae* protected the fly against subsequent exposure to *S. pneumoniae* (Pham, Dionne et al. 2007).

Interestingly, a more recent study found mixed results when testing for a specific primed response in *Drosophila* using diverse combinations of bacteria of varying relatedness to one another. Parasite-specificity on the level of species was shown, but it depended upon which bacterial combination was used for the priming and secondary challenges. In other words, not all pathogens used in the experiment induced a specific primed immune response. Fully reciprocally designed experiments also found no evidence that a higher degree of specificity above the level of bacteria species (i.e. between strains of the same parasite) occurred (Sophie Armitage, personal communication). Here, specific immune priming in *Drosophila* appears to be bacteria and bacteria-combination specific.

1.3.5. PRR-triggered signalling pathways and phagocytic cells may be the effectors of specific memory in invertebrates

In an attempt to characterise the priming response in *Drosophila*, Pham et al (2007) tested loss-of-function mutants for the Toll and Imd pathways and discovered the priming effect was Toll pathway-dependent, but not Imd pathway-dependent. Thus, a specific immune signalling pathway is required for the specific primed response to the gram-positive S. *pneumoniae*. As PRRs are important effectors for triggering signal transduction pathways, one could speculate that one or more PRRs could play an integral part in a specific memory response in invertebrate immunity.

Schneider and co-workers also discovered that by blocking phagocytosis with polystyrene beads in both primed and naïve flies, the protective effect against *S. pneumoniae* disappeared (Pham, Dionne et al. 2007). Specifically, after lethal doses of *S. pneumoniae* were given to primed as well as naïve flies with inhibited phagocytosis, it was discovered that flies of both treatments died at the same rate. This presented indirect evidence that phagocytosis was critical in the primed response in *Drosophila* (Pham, Dionne et al. 2007). Independently, Roth and Kurtz (2009) having developed *in vitro* phagocytosis assays in a fully reciprocal design, directly demonstrated that phagocytic activity in the woodlouse, *Porcellio scaber*, increased in a pathogen strain-specific manner. Using two different strains of heat-killed *Bacillus thuringiensis* bacteria, challenge with homologous combinations (i.e. strain 1 followed by strain 1, or strain 2 followed by strain 2) resulted in significantly higher phagocytic activity than challenge with heterologous combinations (e.g. strain 1 followed by strain 2) (Roth and Kurtz 2009). Thus, phagocytic cells appear to be essential for the specific primed immune response seen in *Drosophila*, and are at least important in the *Porcellio* woodlouse.

Interestingly, similar to the recent work on specific primed immune responses in *Drosophila* (Sophie Armitage; personal communication), both studies found that the characteristics of the specific primed response were not ubiquitous across all bacteria tested. Thus, this further strengthens the evidence against a general induction of the immune system being a cause of the effect. Consequently, it is tempting to speculate that these coarse specificity and memory responses may only have evolved to detect natural pathogens of *Drosophila* and other invertebrates, allowing for some cross-reactivity with those most closely resembling natural pathogens. Furthermore, it has been proposed that "organisms may be selected to only develop acquired immunity to some of the diseases that they encounter" (Boots and Bowers 2004); a proposal supported by both the presumed costliness of evolving and utilising an acquired immune system (Boots and Bowers 2004), together with the known limited

repertoire of germline encoded PRRs available to invertebrates (Janeway and Medzhitov 2002). Nevertheless, invertebrate defences do not appear to lack complexity, but the genetic and cellular mechanisms that underlie specificity and memory in the invertebrate remain to be elucidated.

1.4. Down syndrome cell adhesion molecule

A potential candidate for the phenomenon of specific memory in invertebrates is Down syndrome cell adhesion molecule (Dscam). Dscam is a receptor of the vertebrate and invertebrate nervous systems which has recently been implicated in having a role in the invertebrate immune response (Watson, Puttmann-Holgado et al. 2005). In invertebrates, it is unique in displaying truly exceptional diversity by producing many different isoforms from a single gene through the mechanism of mutually exclusive alternative splicing (Schmucker, Clemens et al. 2000). For example, in *Drosophila melanogaster*, more than 38,000 different isoforms are theoretically possible, and almost all are thought to be expressed and convey equally different interaction specificities. In theory, the provision of hyper immunological diversity for invertebrate immunity could be conferred by *Dscam* and as a consequence the gene could convey a much greater level of pathogen-recognition specificity and even specific memory in invertebrates. Thus, the alternative splicing of *Dscam* in invertebrate immune responses could have important implications for the control of infectious diseases.

Originally, the vertebrate DSCAM was the first of the *Dscam* genes to be isolated and characterised as a neural cell adhesion molecule which mapped to the chromosome band associated with the Down syndrome in humans (Yamakawa, Huo et al. 1998). While looking for candidate genes for the Down syndrome, the researchers discovered that human DSCAM was involved in the development of the nervous system. Soon after, Schmucker et al (2000)

isolated *Dscam* in *Drosophila* and discovered it had a role in axon guidance of the insect's nervous system. Interestingly, while *Drosophila Dscam* exhibits a remarkable level of molecular diversity through alternative splicing (Schmucker, Clemens et al. 2000), the vertebrate DSCAM appears to generate only three transcripts through cryptic splicing sites; indicative of functional divergence of vertebrate and insect *Dscam* (Crayton, Powell et al. 2006). *Dscam* was originally linked to an immune function when an extensive diversity of Dscams was found to be expressed in *Drosophila* haemocytes (Watson, Puttmann-Holgado et al. 2005).

1.4.1. Dscam molecular biology

Dscam is a single-pass membrane receptor and a member of the immunoglobulin superfamily of proteins, a large group of proteins primarily involved in recognition and adhesion (Schmucker, Clemens et al. 2000). One of its functions is to guide axons and dendrites of the nervous system thus ensuring organisation of the intricate nerve cell circuitry (Chen, Kondo et al. 2006; Hattori, Demir et al. 2007; Schmucker 2007). It has a relatively conserved structure across diverse taxa and contains ten immunoglobulin (Ig) domains and six fibronectin type III (FnIII) domains usually in a 9Ig-4FnIII-Ig-2FnIII arrangement, followed by transmembrane and intracellular domains (Hattori, Demir et al. 2007; Shapiro, Love et al. 2007) (but see Chou, Chang et al. 2009). Highly variable exon clusters which express only one of several exon variants in the final mRNA transcript encode part or all of three of the N-terminus-end Ig domains of Dscam. In *Drosophila Dscam*, exon 4 encodes only one alternative version of the first half of Ig2, exon 6 encodes one alternative version of the first half of Ig3, exon 9 encodes alternative versions of the entire Ig7, while exon 17 contributes one of two unique transmembrane domains (Fig.3) (Schmucker, Clemens et al. 2000).

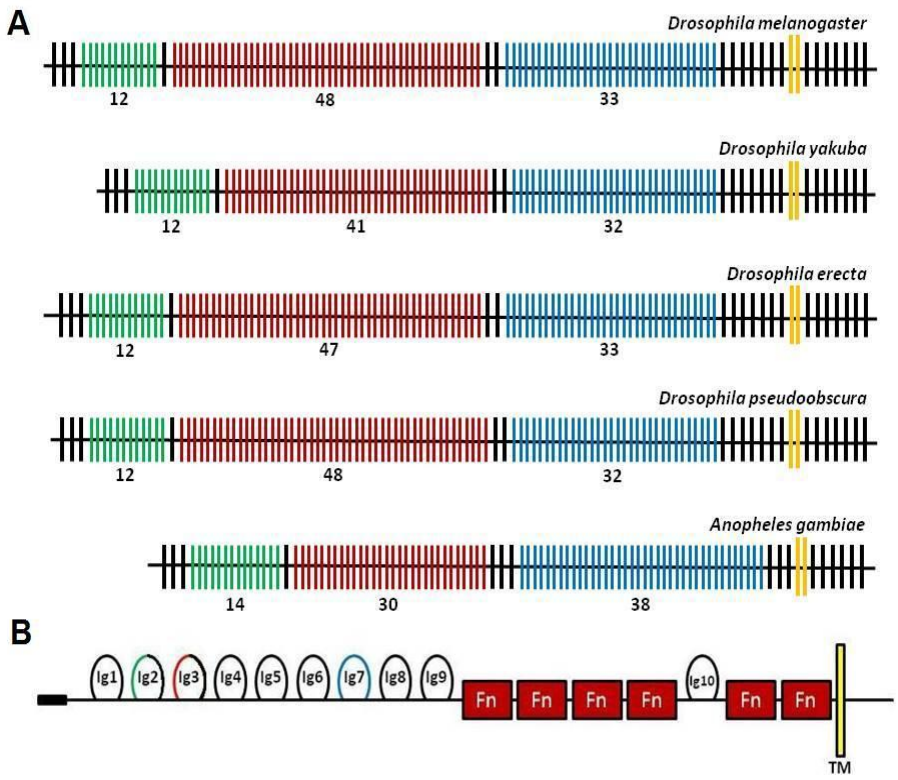


Figure 3: The gene and protein structure of *Drosophila* and *Anopheles Dscam*. (A) *Dscam* gene structure of *Drosophila melanogaster*, *D. yakuba*, *D. erecta*, *D. pseudoobscura* and *Anopheles gambiae*. Vertical black lines represent constitutive exons. Variable exons encoding Ig2, Ig3, Ig7 and the transmembrane domain are coloured green, red, blue and yellow respectively. Mutually exclusive alternative splicing ensures only one variant from each variable exon cluster is present in the final mRNA transcript. (B) The Dscam protein structure. The relatively conserved structure contains ten immunoglobulin (Ig) domains and six fibronectin (Fn) domains. In *Drosophila*, variable exon 4 codes the first half of Ig2 (coloured green), variable exon 6 codes the first half of Ig3 (red), variable exon 9 (or exon 10 in *Anopheles*) codes the entire Ig7 (blue) and variable exon 17 codes one of two alternative transmembrane domains (yellow).

The variable expression of many exon variants gives *Dscam* its ability to generate many unique isoforms. For example, through the alternative splicing of a total of 95 variable exons, *Drosophila melanogaster Dscam* can potentially produce as many as 38,016 different isoforms (Schmucker, Clemens et al. 2000), whereas in comparison, the repertoire of *D. yakuba* can theoretically reach 31,488 isoforms using 87 variable exons (Lee, Kim et al.

2010). In *Anopheles gambiae*, 84 exon variants provide the scope for 31,920 splice forms (Dong, Taylor et al. 2006), while 75 alternatively spliced variants could confer the crayfish with up to 22,272 different Dscam receptors (Watthanasurorot, Jiravanichpaisal et al. 2011). The crustacean *Daphnia magna* has the relatively smaller number of 49 variable exons. Moreover, unlike the insect Dscams characterised so far, which have two alternatively spliced variants coding for the transmembrane domain, *D. magna* Dscam appears to have only one. However, unlike insect Dscams, which have a single cytoplasmic tail, four different cytoplasmic tails have been discovered in this crustacean. As a result, *D. magna* may produce up to 13,056 different protein isoforms (Brites, McTaggart et al. 2008). Thus, the generation of many different Dscam isoforms each with different recognition and binding specificities, could render a system capable of equally unique and diverse interactions (Hattori, Demir et al. 2007; Zinn 2007).

1.4.2. Dscam interactions involve homophilic and heterophilic binding

The tertiary structure of the Dscam protein accommodates two different types of interactions; homophilic and heterophilic. *Homophilic* interactions have been shown to be facilitated by a part of the protein surface known as *epitope I*, and *heterophilic* interactions are thought to occur via a part of the protein surface known as *epitope II* (Fig.4) (Meijers, Puettmann-Holgado et al. 2007).

Epitope I is coded by the 5' portions of the alternative exons 4 and 6, while epitope II is coded by the 3' portion of alternative exon 4 variants and the central portion of alternative exon 6 variants (Meijers, Puettmann-Holgado et al. 2007). Through homophilic interaction with epitope I of one Dscam molecule, and the epitope I of a homolog, the Dscam-Dscam dimer complex is important in cell-to-cell interactions during the development of the nervous system (Wojtowicz, Flanagan et al. 2004; Chen, Kondo et al. 2006; Hattori, Demir et al. 2007; Schmucker 2007). Conversely, through heterophilic interaction, it is thought that

epitope II interacts with molecules other than Dscam homologs, and it has been suggested epitope II could engage in immune responses in invertebrates (discussed further below) (Boehm 2007; Meijers, Puettmann-Holgado et al. 2007; Brites, Encinas-Viso et al. 2011).

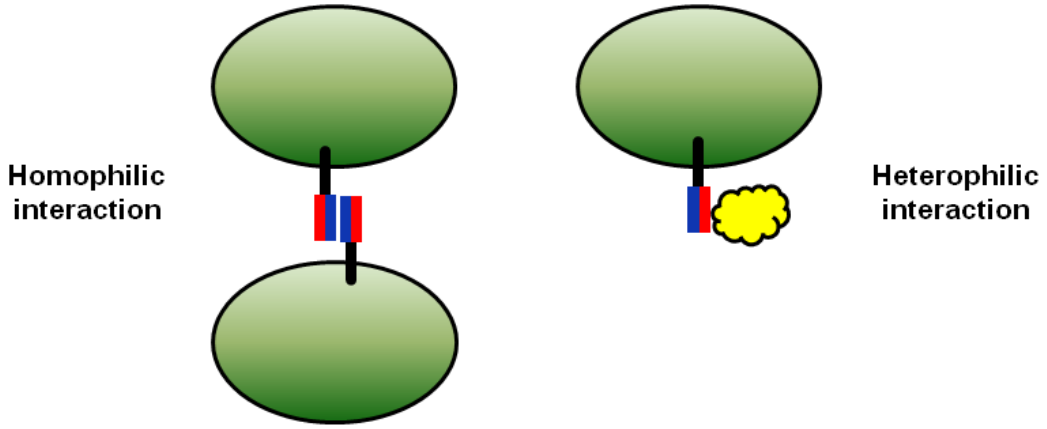


Figure 4: Dscam protein structure accommodates homophilic and heterophilic interactions. Homophilic binding occurs when epitope I (shown in blue) of a Dscam receptor interacts with epitope I of another Dscam molecule. Heterophilic interaction with molecules other than Dscam homologs (shown in yellow) is thought to be facilitated by epitope II (shown in red).

1.4.3. Dscam splicing is complex and unique to each variable exon cluster

The combinatorial splicing of multiple exon variants which complete the final Dscam transcript is important to confer each protein isoform with its unique epitope I and II structure and interaction specificity. Splicing of *Dscam* appears to be complex. It has been shown that each variable exon cluster in *Drosophila Dscam* could potentially have different alternative splicing mechanisms (Graveley 2005). For example, the alternative splicing mechanism for exon 6 uses a docking site and a selector sequence, along with a cluster-specific splicing repressor to ensure that only one alternative exon 6 variant is selected for the Dscam mRNA transcript. The docking site is located within the intron between

conserved exon 5 and exon variant 6.1, while the selector sequences for each exon 6 variant are located in most cases within the short introns downstream of each variant (Fig.5).

In order for a specific exon 6 variant to be included in the final mRNA transcript, the selector sequence must interact with the docking site. Each of the selector sequences is complementary for a part of the docking site, and interaction occurs by the juxtaposition of the selector sequence and the docking site sequence. Importantly, only one selector sequence can interact with the docking site at any one time. This helps ensure mutually exclusive splicing of *Dscam* exon 6 (Graveley 2005). In addition, the splicing repressor has been shown to be crucial for the fidelity of *Dscam* exon 6 splicing also. RNAi depletion of hrp36, a globally acting RNA binding protein, results in Dscam transcripts with the inclusion of multiple exon 6 variants (Olson, Blanchette et al. 2007). Such Dscam transcripts with multiple exon 6 variants do not produce functional proteins (Graveley 2005). Conserved elements similar to the docking site or selector sequences of exon 6 are not readily apparent in either exon 4 or exon 9. Thus, this mechanism of ensuring mutually exclusive alternative splicing appears to be unique to exon 6 only (Graveley 2005). Indeed, a separate study in *Drosophila* tested whether the choice of one exon 9 variant biased the choice of variants from exons 4 and 6. Researchers amplified cDNA with one primer bound in constitutive exon 3, and the other bound in *one* of the alternative variants of exon 9. They discovered that no particular exon 4 or exon 6 variant specifically associated with the selected exon 9 variant, confirming that alternative splicing occurred independently at exons 4, 6 and 9 in *Drosophila* (Neves, Zucker et al. 2004). Thus, it appears that there are different alternative splicing mechanisms for each of the variable exon clusters. At the present time, some information on the regulation of *Drosophila* exon 4 splicing is available (Kreahling and Graveley 2005), but the splicing mechanisms for exons 4 and 9 remain to be more fully characterised, as do those of *Dscams* in other animal species.

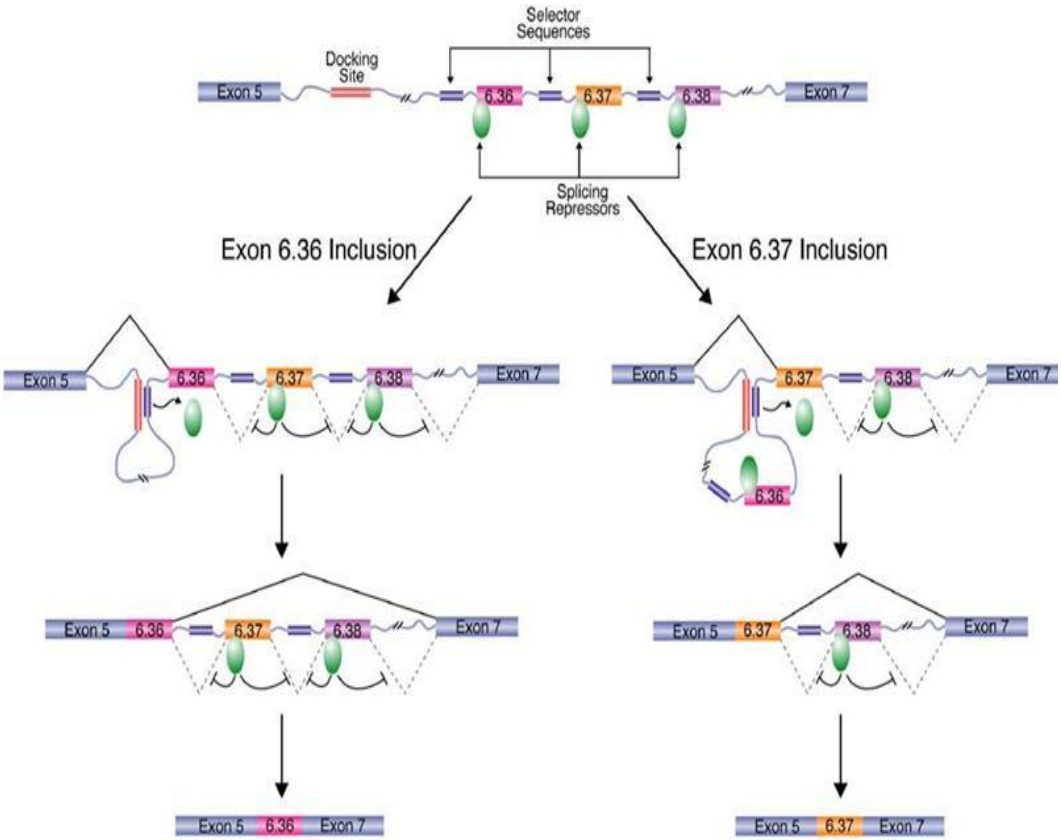


Figure 5: The mechanism of *Dscam* exon 6 mutually exclusive alternative splicing. Key components of *Dscam* exon 6 splicing are the selector sequences which are downstream of each exon 6 variant, and the docking site, located in the intron between conserved exon 5 and variable exon 6.1. Splicing repressors ensure exon 6 variants do not become spliced together, thus ensuring mRNAs do not contain multiple exon 6 variants. Alternative exons 6.36 and 6.37 are used here as examples. For exon 6.36 to be included in the *Dscam* mRNA for instance, the selector sequence downstream of this variant interacts with the docking site by juxtaposition. The repressor is inactivated allowing exon 5 and exon 6.36 to be spliced together. As all other exon 6 variants remain actively repressed, and only one selector sequence can interact with the docking site at any one time, exon 7 is then spliced to the 3' end of exon 6.36, thus ensuring the mutually exclusive splicing of exon 6. Image taken from (Graveley 2005).

1.4.4. Dscam has a role in neuronal development

Dscam operates to direct the movement of sensory neurons in proper nervous-system circuit assembly (Hattori, Demir et al. 2007). More specifically, isoform-specific homophilic binding of Dscam receptors results in the coordination of neighbouring cells or cell processes sharing the same receptive field (Hattori, Demir et al. 2007; Matthews, Kim et al. 2007). This homophilic interaction mediates self-avoidance by triggering downstream signalling cascades which ultimately control the neuron's directionality and interaction with surrounding cells and tissues (Hattori, Millard et al. 2008; Millard and Zipursky 2008). In *Drosophila*, Dscam binds directly to Dock (Schmucker, Clemens et al. 2000). Dock is a protein that functions upstream of Pak, which is involved in intracellular signalling networks controlling cell migration and phagocytosis (Hing, Xiao et al. 1999; Cote and Vuori 2002). Dscam, Dock and Pak, appear to act synergistically to direct neuronal path-finding, and may accomplish this by making changes to the actin-based cytoskeleton of cells (Schmucker, Clemens et al. 2000).

The diversity of splice forms has been shown to be essential for Dscam's role as a neural cell adhesion molecule. For example, a study in *Drosophila* which reduced the Dscam repertoire to a single isoform discovered that the neural circuits of the mutant animals were severely disorganised (Hattori, Demir et al. 2007). It is estimated that each nerve cell expresses between 14 and 50 unique Dscam mRNA molecules (Neves, Zucker et al. 2004). Although the mRNA expression profile doesn't necessarily reflect the Dscam proteins present in a cell at any one time, this observation appears to be evidence for a mechanism for distinguishing cells in *Drosophila*. Chess and co-workers (2004) also discovered that *Drosophila Dscam* splice variant expression was regulated both spatially and temporally. In other words, the alternative splicing of *Dscam* appears to be regulated both during development and also in different tissues. Thus, *Dscam* alternative mRNA splicing and its regulation is hugely

complicated but essential for the integral role of organising signalling networks involved in development.

1.4.5. The origin of Dscam diversity predates the split of the pancrustaceans

The origin of the alternatively splicing form of the *Dscam* gene predates the common ancestral split between insects and crustaceans around 450-500 million years ago (Lee, Kim et al. 2010). A longstanding question surrounding the evolution of *Dscam*'s multiple exon arrays and the mechanism of alternative splicing concerns when and where these features first appeared. While sequence homology has only been found within the vertebrates, nematodes and arthropods, the common ancestor of which lived on the order of a billion years ago, only the arthropods appear to have multiple exon arrays and the capacity for alternative splicing (Graveley, Kaur et al. 2004; Crayton, Powell et al. 2006; Lee, Kim et al. 2010). However, *Dscam* in the deer tick, *Ixodes scapularis*, has no capacity for hyper-diversity: a recent study found no evidence of alternative splicing in *Ixodes* in the regions between the exons orthologous to *Drosophila* exons 3 and 5, 5 and 7, and 8 and 10, suggesting that this feature may be absent in arachnids (Armitage, Freiburg et al. 2012). Furthermore, although searches have been made in many genomes beyond the pancrustaceans, including within the phyla Deuterostomia and Nematoda, homologs of exon 6 have so far only been discovered in the insect and crustacean lineages (Crayton, Powell et al. 2006; Brites, McTaggart et al. 2008). Thus, it appears that the multiple exon arrays and alternative splicing mechanism could have either been lost several times along the lineage leading to the mammals, or originated along the lineage leading to the insects and crustaceans (Crayton, Powell et al. 2006; Brites, McTaggart et al. 2008).

1.4.6. Dscam variable exon clusters may have evolved independently in response to different evolutionary pressures

The *Dscam* gene is thought to have evolved its expansive alternatively splicing ability by staggered homologous recombination (Lee, Kim et al. 2010). Across the Pancrustacea, the number of individual variants within each variable exon cluster is different; indicative of ongoing rapid evolution (Lee, Kim et al. 2010). The exception is the exon 4 cluster, which is older than the ancestor of insects and appears to have changed very little over the past 300 million years (Crayton, Powell et al. 2006; Lee, Kim et al. 2010). Notably, all characterised Drosophilidae *Dscams* have 12 variants of exon 4. However, exons 6 and 9 have different numbers of variants even within families of arthropods, and these clusters have been constructed mainly during the last 300 million years (Lee, Kim et al. 2010). For example in *Drosophila*, the number of exon 6 variants ranges from 41 to 52 across species, while the exon 9 variants range between 29 and 33 variants across species (Lee, Kim et al. 2010). This suggests that the exon 4 cluster is anciently conserved, while exons 6 and 9 are evolving much more quickly. In fact, this pattern is seen across all characterised arthropod *Dscams*, though most strikingly in the insects. Thus, some evolutionary pressure is driving a rapid expansion of the exon 6 and 9 cassettes, but not of the exon 4 cassette, of every lineage of insect. Exons 6 and 9 may have a more important role in the development of the nervous system than exon 4 based on relative tissue-specific expression levels of variants (Watson, Puttmann-Holgado et al. 2005). Thus, the variability of internal exon duplications within exons 6 and 9 may not be surprising. Different evolutionary pressures across the different exon clusters may also reflect the different alternative splicing mechanisms conferred to each cluster (Graveley 2005). Exon 4 has been implicated in having an important role in immunity (Dong, Taylor et al. 2006). The fact that this exon has been relatively quiescent over the past 300 million years suggests that, if so, it may interact with ancient molecular patterns. Also, the nucleotide alignment between the *Drosophila Dscam* variable exons shows much less divergence in exon 4 than in exons 6 and 9 (Lee, Kim et al. 2010). Graveley and co-workers

found that most exon 4 variants were orthologous to exon 4 variants in other species, an observation not seen in exons 6 and 9. Thus, one could speculate that exon 4 codes for a part of the Dscam protein which interacts only with PAMPs which have changed very little over evolutionary time, such as bacterial peptidoglycan. This may also reflect the lack of ubiquitous specificity for bacteria seen in specific primed immune responses (Pham, Dionne et al. 2007; Roth and Kurtz 2009).

Finally, epitopes I and II appear to be under different selective pressures. Epitope II has been suggested as a part of the protein which interacts with parasites (Meijers, Puettmann-Holgado et al. 2007). Brites et al (2011) found that sequences coding for epitope II in *Daphnia Dscam* showed much greater non-synonymous divergence than those coding for epitope I, which is involved in homophilic binding with other Dscams. Interestingly, although sequences coding for epitope II have diverged more than the rest of the gene in both *Drosophila* and *Daphnia*, studies have not found any evidence for the adaptive evolution in either species (Obbard, Welch et al. 2009; Brites, Encinas-Viso et al. 2011). Subsequently, it has been proposed that the elevated non-synonymous diversity may be maintained by some form of balancing selection, or is indicative of the segregating of weakly deleterious alleles, and thus of fewer selective constraints in the supposed pathogen-binding region of Dscam (Brites, Encinas-Viso et al. 2011). The evolution of *Dscam* splicing diversity remains to be fully explained, as does its requirement within the context of innate immunity.

1.4.7. Dscam diversity could have important implications in invertebrate immunity

Alternative splicing, together with other mediators of molecular complexity such as post-translational modification (of which there are over 200 different types known), allows for the significant expansion of coding capacity of the invertebrate genome (Brett, Pospisil et al. 2002). An increased immune recognition capacity and functional diversity could be

accommodated through these mechanisms. Thus, alternative splicing together with post-translational modification could permit a single gene to mediate alternative immune responses via the production of multiple proteins. At present, the importance of the high diversity of Dscam splice forms in the invertebrate immune response remains to be determined. Recently however, some key studies have provided evidence for the role of *Dscam* and Dscam diversity in invertebrate immunity (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011) (discussed in detail below).

1.5. Previous research – Dscam in Immunity

1.5.1. Dscam is implicated in phagocytosis

Dscam has been associated with phagocytosis so far in the fly, the mosquito and the crayfish (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). Schmucker and co-workers, using anti-Dscam antibodies to block *Dscam* function, discovered that haemocyte-specific loss of Dscam substantially lowered phagocytosis of bacteria in *Drosophila* (Watson, Puttmann-Holgado et al. 2005). Dong et al (2006), studying *Dscam* in the *Anopheles* mosquito, also implicated the receptor in the phagocytosis of bacteria. RNA interference-mediated depletion of *Anopheles gambiae Dscam (AgDscam)* decreased the phagocytic activity of an immune-competent mosquito cell line by around 60% compared to controls. Finally, *Dscam* in the crayfish, *Pacifastacus leniusculus*, was associated with pathogen clearance. Recombinant proteins of Dscam isoforms that specifically interacted with bacteria were found to interfere with bacterial binding to haemocytes. This demonstrated that *Dscam* could have an important function in phagocytosis in this crustacean (Watthanasurorot, Jiravanichpaisal et al. 2011). Thus, *Dscam*

appears to play a critical role in the removal of pathogens via phagocytosis across these different classes of arthropods.

1.5.2. Dscam could play a role in the opsonisation of parasites

Phagocytosis and opsonisation act together in the clearance of pathogens (see section 1.2). Immunoprecipitation assays revealed a soluble Dscam protein in the haemolymph of *Drosophila* (Watson, Puttmann-Holgado et al. 2005). In addition, Schmucker and co-workers directly confirmed that *D. melanogaster* S2 cells secreted Dscam splice forms. Soluble Dscam isoforms have also been found in the mosquito (Dong, Taylor et al. 2006), whereas in the shrimp, *Litopenaeus vannamei*, the first reported Dscam lacking a transmembrane domain and cytoplasmic tail was discovered (Chou, Chang et al. 2009). *L. vannamei Dscam* (*LvDscam*) is unique among *Dscams* because a cell-bound protein form has yet to be isolated and described. However, unexpectedly low levels of *LvDscam* were detected in nervous tissues, and thus, may not be involved in neuronal wiring in *L. vannamei*. The authors speculate however that *LvDscam* could retain opsonic characteristics and also play a role in phagocytosis in the shrimp using an integrin binding motif detected between its Ig6 and Ig7 domains. Thus, soluble Dscams are found circulating in the haemolymph of different species of invertebrate, and may play a role in the opsonisation of pathogens.

1.5.3. Dscam exhibits pathogen-specific splice form expression and may bind directly to pathogens

The hyper-diversity of Dscam splice forms may be utilised for specificity in the invertebrate immune response. Dong et al (2006) challenged *Anopheles* cell culture and whole adult mosquitoes with a range of pathogens, and used quantitative RT-PCR to show that *AgDscam* appeared to produce pathogen-specific splice form repertoires. The authors demonstrated that different ‘suites’ of *Dscam* exon 4 variants seemed to be expressed in response to

diverse pathogens including both gram-types of bacteria; strongly suggestive of an infection-responsive alternative splicing of *AgDscam.* Additionally, they showed that *AgDscam* was a 'determinant of resistance' to both bacteria and *Plasmodium* infection by silencing the *Dscam* gene through dsRNA-targeting and monitoring the effects. Subsequently, Dscam receptors appear to recognise and respond to pathogens as evolutionary divergent as bacteria and *Plasmodium*. Söderhäll and co-workers (2011) reported that specific isoforms could also be induced in the crayfish in response to different pathogens. Moreover, using recombinant proteins of the isoforms induced by gram-positive and gram-negative bacteria, they discovered that crayfish Dscam receptors may directly bind to their corresponding bacteria. Finally, it was discovered that Dscam may also bind directly to bacteria in the mosquito after *AgDscam* isoforms were eluted from the surface of the gram-negative *E. coli* after bacterial binding assays (Dong, Taylor et al. 2006). Thus, infection-responsive alternative splicing of *Dscam* could ascribe diversity with specificity to the invertebrate immune response.

1.5.4. Dscam could have multiple roles in the invertebrate immune response

Dscam appears to be an important receptor of the immune response in different invertebrate species, and one could envisage models for *Dscam*'s role in invertebrate immunity. Dscam expressed on the surface of haemocytes could bind directly to pathogens which may trigger conformational changes of the receptor to facilitate interaction with the Dock protein. Alternatively, *Dscam* expression could be connected with the Toll signalling pathway and its response to infection. In addition to the evidence suggesting that the Toll pathway is essential for the specific primed response in *Drosophila* (Pham, Dionne et al. 2007), it has been recently discovered that components of the Toll pathway, namely Rel-1 and Rel-2, regulate *Dscam* alternative splicing in the *Anopheles* mosquito (Yuemei Dong and George Dimopoulos, personal communication of unpublished data). Thus, it is possible that Dscam may either directly or indirectly activate the Toll pathway in response to infection, or

alternatively, components of the Toll pathway may regulate the alternative splicing of *Dscam* retrospective of the first detection of infection.

It is also conceivable that the interaction of cell-bound Dscams with pathogens stimulate the clonal proliferation of those phagocytic cells expressing the responsive Dscam receptors. For instance, the increased phagocytic activity seen in the specific primed immune response of the woodlouse (Roth and Kurtz 2009) may be explained by increased quantities of pathogen-induced Dscam-expressing phagocytes. Moreover, a study in a pancrustacean has shown increased quantities of haemocytes following infection with a natural parasite (Auld, Scholefield et al. 2010), and it is possible that Dscam-mediated responses may be involved. Finally, Dscam could synergistically interact with other components of the innate immune system in as yet uncharacterised signalling cascades to contribute toward specificity and memory in the invertebrate. Needless to say, many questions remain to be answered about *Dscam*'s role in the invertebrate immune response. Does Dscam confer specificity for pathogens or pathogen types? What degree of discriminatory ability exists? What is the function of soluble forms of Dscam? How is the alternative splicing of exon clusters controlled within the context of an immune response? And is the hyper-diversity of Dscam only conferred to the pancrustaceans?

In this thesis I will address the question of the functional importance of Dscam's diversity in immunity, and whether highly-sensitive elicitor-specific splicing occurs in response to diverse pathogens in three different species of the Pancrustacea.

1.6. Aims and objectives

In Chapter 2, I follow up on the work of Dong et al (2006) who reported pathogen-specific splice-form expression of *AgDscam* exon 4 variants in response to diverse pathogens in a mosquito immune-competent cell line. I began by asking whether overall Dscam expression diversity in adult field-caught mosquitoes increased in response to increasing *within*-species parasite diversity using the natural and medically relevant mosquito parasite, *Plasmodium falciparum*. Recording the expression of both exons 4 and 6, I show that Dscam expression diversity significantly increases in parasite-exposed mosquitoes.

In Chapter 3, I set out to develop and adopt a more powerful method for the study of differential Dscam expression in invertebrate immunity. Switching to the more easily handled and better understood model organism, *Drosophila melanogaster*, and having developed a PCR and sequencing assay using Illumina to obtain detailed splicing patterns in collaboration with Dr Darren Obbard, I report the success of the technique by determining characteristics of Dscam variant expression across treatments exposed to diverse pathogens. Moreover, I report the non-random splicing of exons 4 and 6 in *Drosophila*, and a small but detectable effect of pathogen-exposure on exon 4-expression. However, I found no strong patterns of Dscam expression between treatments overall. Thus, I concluded that while the technique appeared to be effective, the use of whole flies may have 'contaminated' possible immune effects with the presupposed higher levels of Dscam present of the nervous system. In Chapter 4, lessons taken from the previous experiment encouraged the investigation of tissue-specific Dscam expression in *Drosophila* in response to well-characterised immune elicitors. Using the same PCR and Illumina sequencing assay as Chapter 3, I found non-random constitutive expression of exons 4 and 6 in all tissues assayed, and detected tissue-specific expression of exon 4 but not exon 6. Furthermore, the small effect of challenge on exon 4-expression was again evident, but the effects did not dominate over tissue-effects.

In Chapter 5, in order to collect data on Dscam expression on a more distantly related pancrustacean, I studied the differential expression of *Dscam* in the crustacean *Daphnia magna* in response to different strains of its natural endoparasite, *Pasteuria ramosa*. I adopted a PCR and Illumina sequencing approach to detect tissue-specific *Dscam* expression in response to *within*-species infection diversity. Here, I show that *Daphnia Dscam* exons 4 and 6 are non-randomly expressed in all tissues assayed, and that a small but detectable effect on exon 4-expression in response to parasite-exposure was also evident.

Finally, in Chapter 6 (General Discussion), I summarise the field, discuss my own work and its implications, and briefly speculate upon the role of Dscam in invertebrate immunity in light of this work and previous research. Finally, I propose future directions for the study of Dscam.

Chapter 2: Alternative splicing of the *Anopheles gambiae Dscam* gene in diverse *Plasmodium falciparum* infections

This chapter has been published as:

Smith P, Mwangi J, Afrane Y, Yan G, Obbard D, Ranford-Cartwright L, Little T: **Alternative splicing of the *Anopheles gambiae* Dscam gene in diverse *Plasmodium falciparum* infections**. Malaria Journal 2011, **10**(1):156. http://www.malariajournal.com/content/10/1/156

It has been written in collaboration with Jonathan Mwangi, Yaw Afrane, Guiyun Yan, Darren Obbard, Lisa Ranford-Cartwright and Tom Little. Journal formatting has been preserved.

2.1. Summary

In invertebrates, including *Anopheles* mosquitoes, Dscam (Down syndrome cell adhesion molecule) appears to be involved in phagocytosis of pathogens, and shows pathogen-specific splice-form expression between divergent pathogen (or parasite) types (e.g. between bacteria and *Plasmodium* or *P. berghei* and *P. falciparum*). Here, we present the first study of Dscam expression in response to genetic diversity within a parasite species. In independent field and laboratory studies, we compared Dscam expression diversity between mosquitoes fed on blood that was free of *P. falciparum* to mosquitoes exposed to either single or mixed genotype infections of *P. falciparum.* We observed significant increases in *Anopheles gambiae* Dscam (AgDscam) receptor diversity in parasite-exposed mosquitoes, but found only weak evidence that AgDscam diversity rises further upon exposure to mixed genotype parasite infections. Finally, we identified a cluster of AgDscam exon 4 variants that become especially common during *Plasmodium* invasion.

2.2. Background

The innate immune system is common to both invertebrates and vertebrates, and although less well studied than the adaptive immune response, is probably responsible for eliminating the majority of infectious organisms. This is achieved through engulfing cells (e.g. phagocytes), antimicrobial compounds (e.g. defensins) and non-specific reactive intermediates such as nitric oxide (Hoffmann 1995; Medzhitov and Janeway 2000; Janeway and Medzhitov 2002; Beutler 2004; Chang and Klotman 2004). The vertebrate adaptive immune system appears to have more complex features: functional antibodies assembled by V-(D)-J joining of gene segments and diversified by somatic hypermutation accommodate an unrivalled resolution in terms of pathogen recognition for the vertebrate adaptive immune response (Murphy, Travers et al. 2007). Invertebrates, lacking antibodies and having only an innate immune system, rely on germline encoded pattern recognition receptors (PRRs) to detect pathogen associated molecular patterns and initiate a response (Medzhitov and Janeway 2000; Akira and Takeda 2004; Christophides, Vlachou et al. 2004; Christensen, Li et al. 2005).

The absence of an equivalent of the vertebrate adaptive immune system has long fostered doubts that the invertebrate immune system could incorporate specificity and/or memory (Klein 1997; Hauton and Smith 2007; Rowley and Powell 2007). However, the absence of the cellular and genetic components of a vertebrate-like anticipatory immune system does not preclude a functional equivalent in invertebrates (Little, Hultmark et al. 2005), and there exists evidence of enhanced secondary responses to homologous infectious challenges (Cooper and Roch 1986; Hartman and Karp 1989; Kurtz and Franz 2003; Little, O'Connor et al. 2003; Sadd, Kleinlogel et al. 2005; Sadd and Schmid-Hempel 2006; Sadd and Schmid-Hempel 2007; Johnson, van Hulten et al. 2008; Roth, Sadd et al. 2009). Moreover, in invertebrates, the genetic background of both hosts and parasites plays a critical role in

determining the probability of infection, a phenomenon called genetic specificity (Carius, Little et al. 2001; Schmid-Hempel and Ebert 2003; Little, Hultmark et al. 2005). Thus invertebrate defences do not lack sophistication, but the genetic and cellular mechanisms that underlie either invertebrate genetic specificity or enhanced secondary responses remain obscure (but see Pham, Dionne et al. 2007).

Alternative splicing could permit a single gene to mediate alternative immune responses via the production of multiple proteins, and the flexibility of such a mechanism could have important implications for the spread of resistance alleles (Harding, Hansen et al. 2005). The Down syndrome cell adhesion molecule (Dscam), which can take some tens of thousands of different forms through alternative splicing, is commonly associated with its function in the vertebrate and invertebrate nervous systems, but seems to also play a role in invertebrate immunity (Du Pasquier 2005; Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Brites, McTaggart et al. 2008). In the fruit fly *Drosophila*, *Dscam* is expressed in cell types that play major roles in the fly's immune system, and RNA interference-mediated depletion of *Dscam* was shown to impair the insect's capacity to engulf bacteria by phagocytosis (Watson, Puttmann-Holgado et al. 2005). Similarly, the silencing of *Anopheles gambiae Dscam* (*AgDscam*) compromises the mosquito's ability to resist *Plasmodium* (Dong, Taylor et al. 2006). Moreover, *AgDscam* produces pathogen-specific splice form repertoires upon immune challenge (Dong, Taylor et al. 2006). In particular, the Dscam repertoire in response to parasite exposure differs between bacteria and *Plasmodium* and between *Plasmodium berghei* and *Plasmodium falciparum* (Dong, Taylor et al. 2006). However, such specificity has so far only been observed in studies comparing these divergent *Plasmodium* parasites, which probably last shared a common ancestor around 55 million years ago (Volkman, Barry et al. 2001).

The response of *AgDscam* transcription to *P. falciparum* diversity (i.e. within-species rather than between-species parasite exposure) may shed light on the resolution of the innate immune system's specificity and dynamics, as well as its limitations. In theory, 31,920 unique splice forms of AgDscam can be generated through the alternative splicing of 84 variable exons contained within three variable exon cassettes (these are exon 4, exon 6, and exon 10) (Dong, Taylor et al. 2006), and which could potentially contribute to a capability to distinguish between different genotypes of *Plasmodium*. This capability would imply a more specific innate immune response than previously supposed. Here, research is described that relates *P. falciparum* genetic diversity to the expression characteristics of the alternatively spliced Dscam receptor in the *An. gambiae* mosquito.

Two independent experiments were performed. The first was a field study that utilized freshly harvested blood from human subjects for which the genetic diversity of naturally acquired *P. falciparum* infections was characterized. The second experiment was based in the laboratory, where mosquitoes were exposed to either single parasite clones or mixtures of clones contained within human red blood cells in culture following an established protocol (Carter, Ranford-Cartwright et al. 1993). AgDscam receptor diversity was studied in two ways. First, a diversity index was calculated based on exon 4 and exon 6 frequencies to assess whether overall AgDscam expression diversity increased under exposure to *P. falciparum* parasites, and if it increased further with greater parasite infection diversity. Second, it was assessed whether particular *Dscam* exon transcripts were associated with infection diversities (in the field study) or particular parasite genotypes (in the laboratory study).

2.3. Methods

2.3.1. Mosquito infection in Kenya

Blood was obtained from primary school students in Iguhu (34°45'E, 0°10'N) in Kakamega district, western Kenya. The predominant malaria vector species in the area is *An. gambiae s.s.* (Githeko and Ndegwa 2001; Minakawa, Sonye et al. 2004). During and shortly after the rainy season, children (5–14 years of age) were screened for gametocytes by thick blood-films stained in Giemsa's stain.

Gametocyte carriers who had >40 gametocytes/µL of blood and who consented to participate in the study were asked to donate 10 mL of blood, which was obtained intravenously by a clinician, and drawn into heparinized tubes. A total of six gametocyte donors were used in this study (two donors per gametocyte-positive infection group). Most of this blood was used for mosquito infections through membrane feeders, with around 50 µL also spotted onto Whatman paper for later DNA extraction using a Chelex-100 isolation technique (Wooden, Kyes et al. 1993). Methods for infecting mosquitoes are described in (Afrane, Little et al. 2008). Drawn blood was immediately centrifuged at 700 x *g*, and the serum discarded and replaced with human AB serum (Cambrex Bio Science, Walkersville, MD, USA). Blood was then placed in warmed membrane feeders. Five to seven-day old *An. gambiae* Kisumu strain mosquitoes were placed in paper cups at a density of 60/cup and allowed to feed on the infected blood for 30 minutes. Mosquitoes used in this experiment were originally obtained near the Kenya Medical Research Institute in Kisumu (Vulule, Beach et al. 1994), but had been bred in an insectary and adapted to feed from a membrane feeder for many years (thus the mosquitoes were unlikely to be highly polymorphic).

A total of twenty four mosquitoes were used in the field study (three mosquitoes per treatment x four infection groups x two independent replicates). Mosquitoes were transferred

to cages post-blood meal, and then placed into RNAlater (Ambion) at 24-hours post-blood meal. Mosquitoes were harvested 24-hours post-feeding because this is the peak time that *Plasmodium* ookinetes penetrate the mosquito midgut (Dong, Taylor et al. 2006). Total RNA extraction was carried out using a Qiagen RNeasy Mini kit. On day 7 after they had been exposed to infected blood, the remaining fed mosquitoes from each cage were dissected in 2% mercurochrome and examined for oocysts to confirm the presence of *Plasmodium* infections.

2.3.2. Plasmodium microsatellite typing

To estimate *P. falciparum* diversity, microsatellite loci were chosen based on their strength in terms of percentage PCR positives, frequency distributions of allele length important for sizing amplicons (Anderson, Su et al. 1999), size in base-pairs (Su and Wellems 1996), and allele frequencies (Anderson, Haubold et al. 2000). A hemi-nested PCR reaction was carried out to amplify six select microsatellite loci from each DNA extraction sample (Anderson, Su et al. 1999). Applied Biosystems Genemapper v4.0 software was used to automate the measurement of allele length and to quantify peaks in samples containing multiple alleles per locus. Only peaks from samples that amplified >200 fluorescent units were included in the deduction of infection diversity. Multiple alleles per locus were scored if the minor microsatellite peaks were >33% the height of the predominant allele.

2.3.3. Mosquito infection in the laboratory

Five to seven-day old female *An. gambiae* (Keele strain) mosquitoes were offered blood meals containing *in vitro* grown gametocytes of two different genotypes of *P. falciparum* (clone 3D7 (Walliker, Quakyi et al. 1987) and clone HB3 (Bhasin and Trager 1984)) through membrane feeders, and three whole mosquitoes per treatment were harvested after 24-hours. Three independent experiments on three different dates were performed. A total of thirty six mosquitoes were used in this study (three mosquitoes per treatment x four treatment groups x

three independent replicates). The gametocyte culture and membrane feeding protocols followed those previously described (Carter, Ranford-Cartwright et al. 1993). The isolation of total RNA was carried out using a Qiagen RNeasy Mini technology kit.

2.3.4. Quantifying Dscam diversity

RNA extracted from three individual mosquitoes per treatment was reverse transcribed using random hexamers and primers were designed to amplify a fragment of *AgDscam* spanning from exon 3 to exon 7 (primer sequences were: (F) 5' - GTATACGCCTGCATGGCTAAGA - 3', (R) 5' - GCCCTTATCCTCCTTCTTG - 3'). Thus, the amplicons comprised variable exons 4 and 6, and the conserved exon 5. PCR products were cloned using a TOPO TA cloning kit with pCR®4-TOPO® vector and transformed in chemically competent *Escherichia coli*. Around 50 clones per mosquito were sequenced in a 96-capillary ABI 3730*xl* DNA Analyzer (provided by the GenePool Sequencing Facility, University of Edinburgh). Sequences were aligned using BioEdit version 7.0.9.0 and identified by cross-referencing with the known *An. gambiae* genome sequence available from the Ensembl genome browser (http://www.ensembl.org/index.html).

2.3.5. Statistical analyses

Diversity was measured with Simpson's Index (1-D) (Simpson 1949), which quantified the combination of expressed exon 4 and exon 6 variants in each transcript as determined by sequencing. The presence and abundance of each individual exon, and exon 4-6 combination was identified. Statistical analysis of data using general linear modelling and one-way analysis of variance was carried out using Minitab 15.1.1.0 software. For the laboratory study, 'clone' was the single fixed effect with three levels, 3D7, HB3 or mixture of both, and for the field study, the number of genotypes detected (single, double or triple infection) was a fixed effect and donor was added as a random effect to account for variation. Data were normalized with a square root transformation of Simpson's Index (D) before analyses. The

Neighbour-Joining trees (see appendices) were produced using PHYLIP and variation in counts of exon groupings was analysed using Pearson's chi-square.

2.4. Results and Discussion

For the field study, *An. gambiae* were membrane-fed on blood samples taken from gametocyte-carrying children, and RNA was harvested from the insects 24-hours post-exposure, a time when parasites are traversing the midgut epithelium (Dong, Taylor et al. 2006). cDNAs were then cloned and sequenced to identify specific *AgDscam* gene variants (at exons 4 and 6) expressed within insects fed from different blood samples. Control blood was taken from children carrying no *Plasmodium* (as detected with microscopy). A set of *P. falciparum* microsatellite markers (Anderson, Su et al. 1999) were used to identify parasite genotypes in gametocyte carriers. As blood samples contain whole populations of parasites, it is not possible to precisely estimate the number of genotypes present. However, by simply counting alleles at each locus, it is possible to identify the minimum number of genotypes present in a sample, and thus the methods used yielded a lower-bound of parasite genetic diversity. Single, double and triple infections were subsequently identified.

AgDscam expression, as characterized by a diversity index (Simpson 1949) and averaged over both studied exons, was affected by the blood that the mosquitoes fed upon in the field ($F_{3,20} = 3.22$, $P = 0.045$; Fig. 1). It appears that AgDscam is more diverse when parasites are present (diversity was significantly lower in uninfected blood controls than in all other samples; Fig. 1). It is the combined diversity at exons 4 and 6 that drive this pattern, as a relationship between expression diversity and parasite diversity was not apparent when the exons were studied separately (exon 4: $F_{1,22} = 3.33$, $P = 0.081$; exon 6: $F_{1,22} = 0.02$, $P = 0.901$).

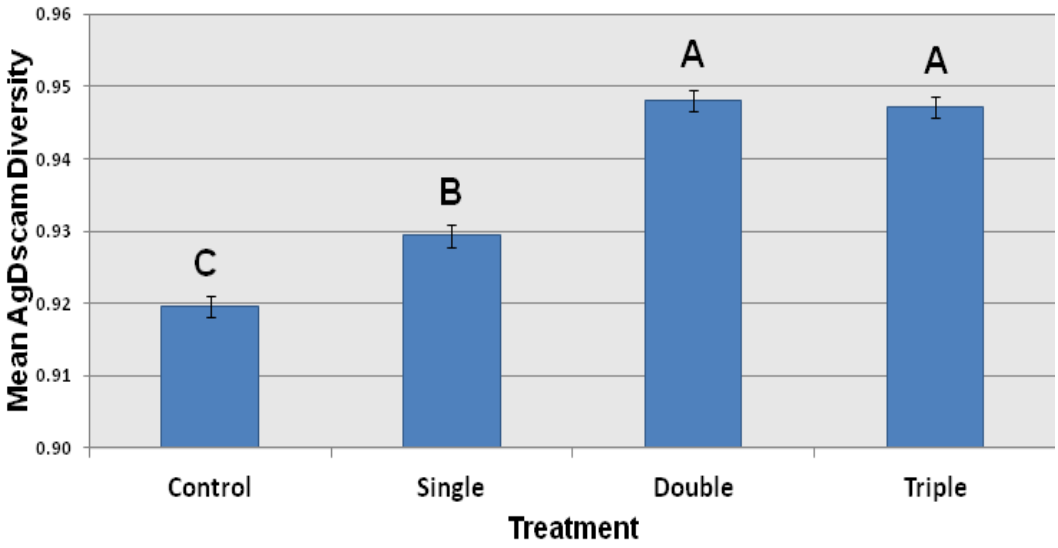


Figure 1: ***Anopheles gambiae Dscam*** **expression diversity versus increasing parasite diversity in the field**. AgDscam diversity (as 1- *Simpson's Index*) increased between controls and exposed treatments and between single and double genotype-exposures only. Levels not connected by same letter are significantly different. Error bars represent standard error (SE). **Key**: Control; blood meal with no parasites, Single; mosquitoes exposed to a minimum of one *P. falciparum* genotype, Double; mosquitoes exposed to a minimum of two *P. falciparum* genotypes, Triple; mosquitoes exposed to a minimum of three *P. falciparum* genotypes.

It was interesting to note that there was a non-statistically significant trend for an association between exon 4 (but not exon 6) diversity and parasite diversity. This may imply that exon 4 has a bigger role than exon 6 in responding to *Plasmodium* in the field. Additionally, AgDscam diversity was also higher when mosquitoes fed upon blood with double infections when compared to single infections, but diversity did not rise further with triple infections (Fig. 1). Thus, these field data provided only limited evidence of a link between host response recognition capacity and parasite intraspecific diversity. As the blood stage diversity as measured by microsatellites could not be certain to be representative of gametocyte diversity, i.e. it is conceivable that not all parasite types were producing gametocytes (see Nwakanma, Kheir et al. 2008), a laboratory-based study was carried out

using gametocyte-producing lines of *P. falciparum*. In this way, the diversity of gametocytes entering the experimental mosquitoes could be certain.

For the laboratory study, mosquitoes were membrane-fed on blood infected with gametocytes of either *P. falciparum* clone 3D7 (Walliker, Quakyi et al. 1987) or HB3 (Bhasin and Trager 1984), or a mixture of the two, and RNA was harvested from the insects 24-hours post-exposure. cDNAs were cloned and sequenced and *AgDscam* exon 4 and exon 6 variants expressed were identified within mosquitoes exposed to different treatments (control blood with no *Plasmodium*, clone 3D7, clone HB3, and clones 3D7 and HB3 mixed). It was found that *AgDscam* expression was affected by the blood that the mosquitoes fed upon ($F_{3,32} = 5.29$, $P = 0.004$; Fig. 2).

Thus, as with the field study, it appears that AgDscam is more diverse when parasites are present (diversity was significantly lower in uninfected blood controls than in all other samples; Fig. 2). Although diversity increased under exposure to either laboratory parasite clone, it did not increase further in response to a mixture of the two clones (Fig. 2). As with the field data, it was the combined diversity at exons 4 and 6 that drove this pattern, as variation in Dscam diversity was not apparent when the exons were studied separately (exon 4: $F_{1,34} = 3.10$, $P = 0.087$; exon 6: $F_{1,34} = 0.16$, $P = 0.690$). Again, there was some evidence of a trend for exon 4 but not exon 6, as previously seen in the field study, implying that the variability rendered by exon 4 in responding to *Plasmodium* may be relatively more important than the variability offered by exon 6-expression.

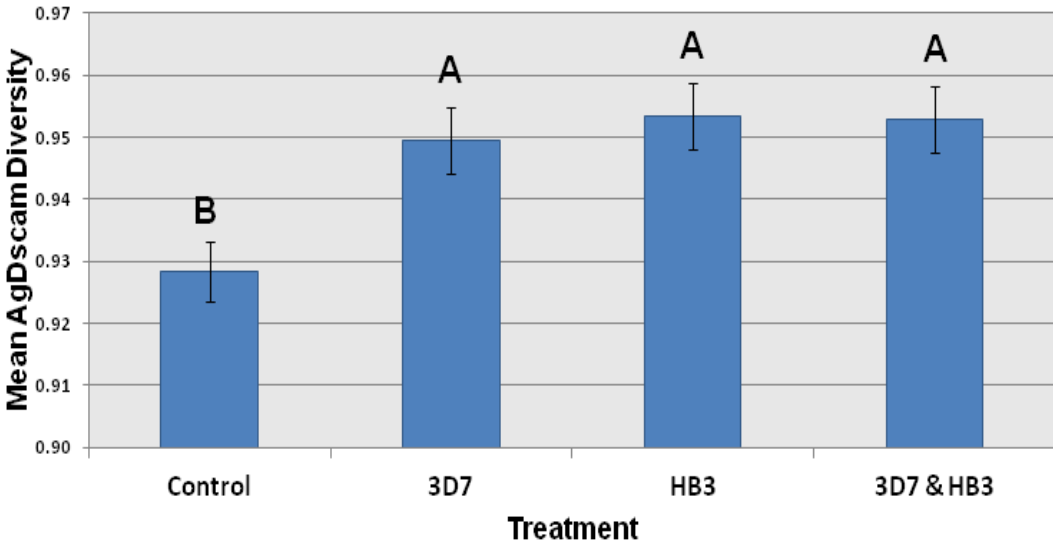


Figure 2: *Anopheles gambiae Dscam* expression diversity versus increasing parasite diversity in the lab. AgDscam diversity (as 1- *Simpson's Index*) increased between controls and exposed treatments, but did not increase between the single genotype-exposure and the mixed genotype-exposure. Levels not connected by same letter are significantly different. Error bars represent standard error (SE). **Key**: Control; blood meal with no parasites, 3D7; mosquitoes exposed to *Plasmodium falciparum* clone 3D7, HB3; mosquitoes exposed to *Plasmodium falciparum* clone HB3, 3D7 & HB3; mosquitoes exposed to both clones in the same blood meal.

A χ^2 analysis was used to identify whether particular splice variants were over- or under-represented in any of the field (control, single, double or triple) or laboratory (control, clone 3D7, clone HB3, mixed) infections. Because the extreme splice variation leads to very low counts for individual exons, exon variants were grouped into categories based on their genetic distance. Although this grouping may not reflect any relationship between biological function and antigen recognition similarity, varying the cut-off points did not change the qualitative results. Based on a Neighbour-Joining (NJ) tree (exon 4, see Fig. S1; exon 6, see Fig. S2), genetic groupings of exon variants were defined such that exon 4 variants were clustered into 3 groups (Fig. S1), and exon 6 into 3 groups also (Fig. S2).

χ^2 contingency tests were used to determine whether the number of observations (counts) of each group differed between the infection categories (see appendix for raw data). In general, no exons were over- or under-represented in any particular infection grouping in either the field or laboratory experiments. The only exceptions to this were exon 4 variants 4.11, 4.12 and 4.13 which are clearly a distinct genetic grouping (see Fig. S1), and were under-represented in control mosquitoes compared to exposed treatments in the field (Pearson's Chi-Square = 6.318, DF = 2, P = 0.042), and exon 6 variants 6.1 and 6.9 which are also a distinct genetic grouping (see Fig. S2), and were under-represented in control mosquitoes, but also over-represented in mosquitoes exposed to a single genotype of *P. falciparum* in the field (Pearson Chi-Square = 19.975, DF = 6, P-Value = 0.003). These data suggests that exon 4 variants 4.11, 4.12 and 4.13, and exon 6 variants 6.1 and 6.9, may be particularly important for the mosquito's immune response to *Plasmodium* in the field.

2.5. Conclusion

The results show an increase in AgDscam splice-form diversity at 24-hours post-exposure, when the parasites are crossing the insect's midgut epithelium (Dimopoulos, Seeley et al. 1998; Dong and Dimopoulos 2006; Dong, Taylor et al. 2006). This observation, confirmed both in the field and the laboratory, reinforce that the AgDscam receptor responds to *Plasmodium* at a vital stage for the parasite's development. The limited association between AgDscam diversity and *P. falciparum* genotype-diversity in the field raises the possibility that the alternatively-spliced receptor could be responding to *P. falciparum* diversity. This observation, however, was not seen between double- and triple-exposed treatments. Although it cannot be assumed that every parasite genotype detected in the blood samples in the field is represented in the sexual stage, there is an apparent consistency in the results showing a lack of genotype-specific *Dscam* expression diversity following two different experimental approaches (field and laboratory). Naturally, the field data may also be affected

by unmeasured factors, for example some of the blood samples could conceivably have harboured other parasite species. This possibility was investigated using PCR detection for other *Plasmodium* parasites (specifically *Plasmodium malariae, Plasmodium ovale* and *Plasmodium vivax*), and the presence of *P. malariae* in both of the blood samples containing a single genotype of *P. falciparum* was recorded (see appendix, Table S3). Consequently, the single infections were confounded with multi-species infections, and yet AgDscam diversity of these multi-species exposures was lower than the diversity levels of the double and triple *P. falciparum* exposures. Although even higher *Dscam* expression diversity could be expected in this instance, this was not observed, and thus other interpretations are possible depending on whether one species has a stronger influence on *Dscam* expression within a particular multiple-infection than another. It was also interesting to note that the over-representation of the group of exon 6 variants (variants 6.1 and 6.9) found in these multi-species infections was not found in the single-species infections, implying that these variants could be influenced by the presence of *P. malariae*. The significance of these observations remains to be investigated. It was also determined whether *P. falciparum* infection intensity in the blood samples, i.e. abundance of parasite stages (as determined by quantitative PCR), could have affected our results, but we found no relationship between parasite abundance and Dscam diversity (see appendix, figure S3). Finally, future studies may benefit from the use of multiple mosquito genotypes. It is possible that different genotypes may respond differently to the same parasite challenge in terms of *AgDscam* expression. This may be a logical direction for future work as the relative absence of related studies on AgDscam allows little speculation on whether the colonies used in this study would be a good proxy for what can be expected in natural populations. Thus, in summary, the data clearly indicate that AgDscam diversity increases with parasite exposure, but they do not suggest that AgDscam diversity rises further in response to increased parasite diversity.

Chapter 3: Dscam gene expression in response to diverse pathogens in the fruit fly *Drosophila melanogaster*

This chapter was written with comments from Darren Obbard and Tom Little. Statistical analyses were done in collaboration with Darren Obbard. The quality trimming, counting and sorting of raw Illumina sequencing reads was done using a custom R pipeline written by Darren Obbard. Illumina library construction, sequencing and base-calling were performed by the GenePool Sequencing Facility (University of Edinburgh).

3.1. Introduction

Evidence of enhanced immunity and pathogen-specific memory has been reported in several members of the pancrustaceans (see Chapter 1, General Introduction). For example in *Drosophila*, a specific primed response was found after flies were injected with a sub-lethal dose of *Streptococcus pneumoniae* which conferred protection from a subsequent lethal dose of the same gram-positive bacterium, but did not protect against other pathogens (Pham, Dionne et al. 2007). The specific priming effect was dependent upon phagocytes, immune cells required for the removal of pathogens (Lemaitre and Hoffmann 2007), and the Toll signalling pathway, which responds to the presence of gram-positive bacteria and fungi (Akira and Takeda 2004).

A candidate gene for the specific primed response in *Drosophila* is Down syndrome cell adhesion molecule (Dscam). Dscam is an axon guidance receptor that displays exceptional diversity of isoforms through the mutually exclusive alternative splicing of many different gene variants. For example, in *Drosophila melanogaster*, 95 variable exons contained within 3 exon cassettes can produce up to 38,016 different proteins with different interaction specificities (Schmucker, Clemens et al. 2000). This diversity has been shown to be essential for *Dscam*'s role in the fly's nervous system (Hattori, Demir et al. 2007), and evidence is emerging that suggests the alternative splicing gene may also be important in invertebrate immunity. In other species, specific *Dscam* exon variants have been reported to be over-expressed in response to particular pathogens and pathogen-types (Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). Additionally, *Dscam* has also been implicated in phagocytosis and may be connected with the Toll signalling cascade in response to infection. Specifically, the silencing of *Dscam* results in significant depletion of phagocytic activity in both the fly and the mosquito (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006), while pathogen-specific Dscam recombinant proteins interfere with

bacterial binding by phagocytic cells in the crayfish (Watthanasurorot, Jiravanichpaisal et al. 2011). Furthermore, components of the Toll signalling pathway have been shown to regulate Dscam alternative splicing in the mosquito (Yuemei Dong and George Dimopoulos, personal communication of unpublished data). Thus, the capacity for high expression diversity and recognition specificity together with an implicated role in phagocytosis and possible involvement with the Toll signalling pathway, make Dscam a candidate for specific recognition in pathogen-clearance in *Drosophila*.

In Chapter 2, I set out to determine whether overall Dscam expression diversity increased in the adult *Anopheles gambiae* mosquito in response to its natural parasite, *Plasmodium falciparum*. The data suggested that Dscam expression diversity increased in response to the presence of parasites in both the lab and the field, but showed only weak evidence of a further increase in response to increasing parasite diversity. Thus, while *Anopheles Dscam* appeared to play a role in the mosquito immune response to parasites, the phenomenon of parasite genotype-specific splicing in adult mosquitoes could not be confirmed. However, 'diversity' was a relatively uninformative measure in this case as the diversity of exon variant expression is very high in *Dscam*, and subsequently the sensitivity of detection was low. In other words, the cloning and sequencing approach carried out in Chapter 2 could not give enough power to enable a good estimate of splice-variant frequencies as there are so many of them. To quantify the relative frequencies of more than 400 different entities (the number of possible exon 4-6 combinations of *Anopheles Dscam*) by making a total of just 50 observations per insect is unrealistic. As such, a more desirable measure would be a quantification of the Dscam expression pattern. Thus, methods which could permit a greater sampling of Dscam transcripts would be preferred.

Here, I infected an invertebrate host with several pathogens and then used qRT-PCR assays to determine whether immune responses occurred, and also to help determine whether the

chosen post-infection harvest time was appropriate to study Dscam expression (i.e. during heightened immune responses). I then used a new Illumina-based sequencing protocol to identify expression levels of Dscam exon variants, and carried out an independent cloning and sequencing assay to compare the output of the different sequencing methods. I also compared the Illumina sequencing results to a past microarray study to establish whether they were consistent. I used two main complementary approaches to the statistical analysis. A generalised linear mixed model implemented in MCMCglmm (Hadfield 2009) was used to infer the effect of immune challenge on Dscam expression patterns, and the widely used DESeq (Anders and Huber 2010) analysis was implemented to study differential expression of individual *Dscam* exon variants between treatments.

The remainder of this introduction is dedicated to (1) describing the assays currently used to study gene expression and alternative splicing of *Dscam*, (2) to outline a new high-throughput sequencing approach designed to study alternative splicing of *Dscam* in the present study, (3) to introduce the organisms used in this work, and (4) to outline the aims of the study.

Assays currently used for the study of gene expression and alternative splicing in Dscam

Cloning and Sequencing

The sequencing of cloned PCR products has been widely used to study *Dscam* gene expression. The differential use of alternative *Dscam* exons was first detected in *Drosophila* by sequencing across the variable and constitutive exons of the gene (Schmucker, Clemens et al. 2000), and different combinations of exon variants spliced together in the same transcripts were found in the olfactory receptor neurons of fruit fly antennae by sequencing cloned Dscam cDNAs (Hummel, Vasconcelos et al. 2003). Watson et al (2005) cloned cDNAs in *Drosophila* spanning from constitutive exon 3 to constitutive exon 7 and

discovered that Dscam mRNAs in different tissues contained many different expressed exons and exon 4-6 combinations (Watson, Puttmann-Holgado et al. 2005). Later, Brites et al (2008) sequenced clones of Dscam cDNAs in *Daphnia magna* to obtain the first Dscam sequences from this crustacean and also identify associations between different exon variants in individual Dscam transcripts. Finally, Chou et al (2010) cloned and sequenced partial cDNA fragments from the shrimp *Litopenaeus vannamei*, to also identify individual and combinatorial expression of Dscam exon variants.

The cloning and sequencing assay employed in Chapter 2 of this thesis was effective in determining the expression of exon 4 and 6 variants independently, and the specific exon 4-6 combinations. As such, this method provided both quantitative information (counts of individual exon variants) and qualitative information (combinations of exons) about *Dscam* gene expression in a natural host-parasite system. However, to permit more robust conclusions from cloning and sequencing, extensive sequencing of a statistically large number of randomly selected cDNAs is essential (Hall 2007). In *Drosophila* for instance, *Dscam* exon 4-6 combination-diversity alone extends up to 576 possible isoforms, and my previous work in *Anopheles* (see Chapter 2) found the distribution of exon combination-expression to be in the direction of even, indicating that all isoforms may be rare. The sampling of high numbers of Dscam transcripts using the cloning and sequencing method would be impractical. Thus, to properly investigate Dscam's putative pathogen-specificity, a method enabling the sampling of a much larger amount of Dscam mRNAs is needed.

Single-strand conformation polymorphism gel electrophoresis

Single-strand conformation polymorphism (SSCP) gel electrophoresis has also been used to analyse alternative splicing in *Dscam* (Celotto and Graveley 2001). Conformational differences between single-stranded nucleotide sequences of the same length can be distinguished on a gel, and is the basis of SSCP assays (Orita, Suzuki et al. 1989). Celotto

and Graveley (2001) used SSCP gel electrophoresis to distinguish most of the 12 different exon 4 variants of *D. melanogaster* from RT-PCR products. As all *D. melanogaster* exon 4 variants are similarly sized (ranging between 159 and 171 base pairs) the identification of individual variants using traditional agarose gel electrophoresis would not be possible (Celotto and Graveley 2001). However, a drawback of SSCPs is that some *Dscam* exon 4 variants co-migrate on gels (Celotto and Graveley 2001). As such, the relative frequency of most but not all exon 4 variants expressed in each RNA sample can be determined. This problem may be elevated in exon clusters which contain even more variants such as the *D. melanogaster* exons 6 and 9, which contain 48 and 33 mutually exclusive splice variants, respectively.

Quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) has been used to detect relative expression levels of *Dscam* exon variants. Dong et al (2006) used qRT-PCR to study the relative expression levels of *Dscam* exon 4 variants in response to immune elicitors in *Anopheles*, and subsequently discovered that some exon 4 variants were over-represented in response to specific pathogens. Moreover, the abundance of Dscam transcripts and the specific usage of exon 4 variants in response to parasites were quantified in the crustacean *D. magna* using this method (Brites 2010). The study of *Dscam* exon-exon combinations using qRT-PCR has not been attempted to my knowledge, and it would not only require primers for all exon-exon junctions to determine exon to exon splicing, but would also require the amplification of fragment sizes much larger than would normally be considered optimal for high primer efficiency. For example, to detect the splicing of exon 4 and exon 6 in *D. melanogaster*, fragment sizes of over 200bp would need to be amplified, more than the preferred 75-150bp fragments, while substantially larger fragments would be required to determine the splicing of variable exon 6 with variable exon 9. Furthermore, similarities between exon variant sequences would make the design of efficient primers difficult.

Microarrays

Microarray technology has been used to measure the total level of expression of Dscam gene variants in *Drosophila*. Using microarrays, Watson et al (2005) detected a high diversity of isoform expression in different tissues of the insect (Watson, Puttmann-Holgado et al. 2005). Although microarrays can simultaneously detect the expression intensities of *Dscam*'s many gene variants individually, the splicing together of specific exon 4-6 combinations is missed, and would require the design of complex splice-sensitive arrays. Detecting the combinatorial splicing of *Dscam* using microarrays would require probes to be designed that are specific to each splice-junction. Prior knowledge of exon-exon junctions are required before splice probes can be designed, and therefore this method may miss previously unknown exon splicing events and also previously unknown allelic variation within alternative exons (Lee and Roy 2004). Moreover, as with a qRT-PCR approach, splice probes for each exon-exon junction may be difficult to design due to similarities between alternative exons (see Srinivasan, Shiue et al. 2005), which could result in cross-hybridization of probes and ultimately affect measures of exon variant expression. Thus, although microarray technology has undoubtedly been a powerful tool in the study of *Dscam*, the study of combinatorial splicing and its possible implications would require a more practical and powerful method.

The development of a new sequencing assay adopting Illumina technology

The arrival of high-throughput sequencing and its ability to provide hundreds of thousands of counts for a specific expressed gene or gene variant, permits a considerably greater sampling of a cDNA pool (Mardis 2008). As such, a truer representation of *Dscam* gene variant expression and *Dscam* alternative splicing within the backgrounds of diverse immune challenges could be acquired and may provide more evidence for its role within the context of insect immunity. To help achieve this, we have developed a new high-throughput

sequencing assay using Illumina technology to enable the counting of many tens of thousands of *Dscam* exon variants.

Illumina sequencing technology is one of several new powerful methods for sequencing on a genome-wide scale (Mardis 2008; Schuster 2008; Shendure and Ji 2008; Peng and Zhang 2009). Illumina can yield tens of millions of reads per instrument-run using a massively parallel sequencing process. DNA templates are immobilised on a flow cell, and are then clonally amplified through a series of extensions and bridge amplifications generating many clusters containing millions of copies of the original copy. These clusters are then sequenced simultaneously base by base using four fluorescently-labelled nucleotides which naturally compete, minimising incorporation bias. After each sequencing cycle the clusters are excited by a laser. This emits a colour identifying the newly added base, which is called by measuring the signal intensity (Mardis 2008; Shendure and Ji 2008). Here, we have adopted Illumina technology in the development of a sequencing assay to study *Dscam* alternative splicing. I used PCR to amplify a fragment of *Dscam* spanning from the 3' end of constitutive exon 3 to the 5' end of constitutive exon 7, thus producing amplicons containing complete variable exons 4 and 6, and the constitutive exon 5. For each treatment in the experiment a unique 5bp Molecular ID (MID) tag was incorporated into the 5' PCR primer allowing multiplexed reads to be assigned to their treatment (Fig. 1). Thus, many thousands of Dscam cDNAs were sequenced in this study, and the expression of exons 4 and 6, and the combination thereof, was identified for each treatment.

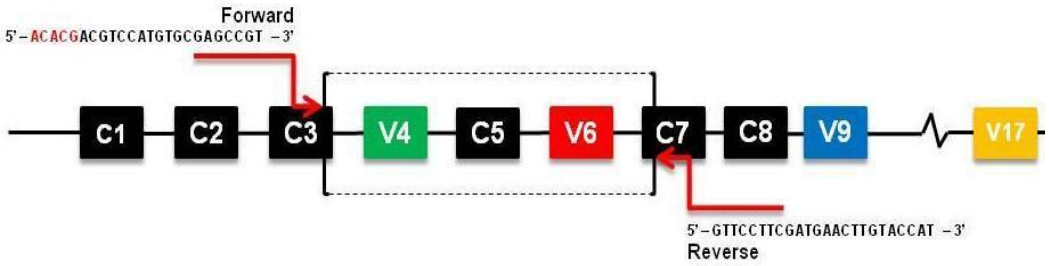


Figure 1: *Dscam* variable exons 4 and 6 were PCR-amplified simultaneously and MID tags were incorporated into the forward primers. *D. melanogaster* Dscam amplicons comprised variable exons 4 and 6, together with constitutive exon 5. Unique molecular ID (MID) tags (here with ACACG used as one example, shown in red) were incorporated into forward primers for each treatment in the study. This allowed the identification of treatments in multiplexed sequencing reactions. Prepared samples were then sequenced using Illumina technology and the expression of exon 4, exon 6, and the exon 4-6 combination for each cDNA was recorded. Primer sequences shown are those used in the experiment. Constitutive exons are represented by black squares and are prefixed with the letter 'C', and variable exons are represented by coloured squares and are prefixed with the letter 'V'. All exons are numbered as they appear in the transcript.

Drosophila immune responses as a model

In this study, I challenged *Drosophila melanogaster* with an array of diverse immune elicitors. I aimed to expose flies to immune challenges which would provide the widest possible comparisons and as such, I chose two different bacteria (one gram-positive and one gram-negative), two different viruses (RNA and DNA), and a fungus.

The gram-negative bacteria, *Escherichia coli*, triggers the up-regulation of the anti-bacterial peptide Diptericin in *D. melanogaster*, while the gram-positive bacterium, *Micrococcus luteus*, has been shown to up-regulate the anti-microbial compound Drosomycin (Lemaitre, Reichhart et al. 1997). *Drosophila* C virus (DCV) is a single-stranded RNA virus and has

been found to infect *D. melanogaster* and *D. simulans* in the wild (Jousset, Bergoin et al. 1977; Kapun, Nolte et al. 2010). DCV infection up-regulates the expression of the Jak/STAT signalling pathway effector gene, vir-1 (Dostert, Jouanguy et al. 2005; Hedges and Johnson 2008), and DCV induction of vir-1 is regarded as an important model for viral infection in *Drosophila* (Ip 2005). Flies were also exposed to *Plodia interpunctella* granulosis virus (PiGV). PiGV is a DNA virus of the Indian meal moth, *Plodia interpunctella*. No DNA virus infecting *Drosophila* was known when these experiments were performed, thus an aim of introducing PiGV into the fly was to attempt to provoke an atypical immune response. Finally, the fungus *Beauveria bassiana* is an entomopathogenic fungus found in soil that can parasitize various arthropod species (Tinsley, Blanford et al. 2006) and is naturally pathogenic to *Drosophila*. It has been shown to up-regulate Drosomycin in *D. melanogaster* (Lemaitre, Reichhart et al. 1997).

Many *Drosophila* species including *D. melanogaster* are naturally infected with the endosymbiont *Wolbachia*, a cytoplasmically-inherited bacterium that appears to mediate virus protection (Hedges, Brownlie et al. 2008; Teixeira, Ferreira et al. 2008; Osborne, Leong et al. 2009). For example, after clearing *Wolbachia* infection in *D. melanogaster* with tetracycline, Teixeira and co-workers showed that both DCV viral load and viral sensitivity increased in the fly (Teixeira, Ferreira et al. 2008). This effect was seen for other RNA viruses but not for a DNA virus. In contrast, *Wolbachia* infection does not appear to confer antibacterial protection in *D. melanogaster* (Wong, Hedges et al. 2011). Thus, the presence or absence of *Wolbachia* in *D. melanogaster* may confer different outcomes to virus infections. Consequently, I also used *Drosophila* both naturally infected with *Wolbachia*, and cleared of *Wolbachia* using antibiotics, to investigate any detectable differences in Dscam expression between the two infection states in response to DCV infection.

Aims

Here, a new Illumina sequencing assay was employed to investigate the expression of the *Drosophila Dscam* gene in whole flies in response to diverse immune challenges including bacteria, viruses, and a fungus. I aimed to (1) investigate whether the deep sequencing assay could detect differences in exon variant expression, detect novel exons, and identify unexpressed exons, (2) investigate whether it could detect combinatorial splicing in *Drosophila* and determine whether this splicing is random or non-random, and (3) determine whether infection-responsive alternative splicing occurs in response to a wide array of immune elicitors, including in flies harbouring a putatively protective endosymbiotic bacterium.

3.2. Methods

3.2.1. Fly rearing

Flies used in the experiment were the *D. melanogaster* Oregon R line (original stock provided by Brian Charlesworth, University of Edinburgh) and maintained in the lab on standard fly medium (1000ml water, 10.3g agar, 140.6g sugar, 103g maize, 28.1g yeast and 22.5ml methylparaben; an antifungal and antibacterial agent) at 18°C or 25°C with a 12:12h light:dark cycle. Flies were treated with 0.03% tetracycline hydrochloride for three generations to clear *Wolbachia* intracellular bacteria for *Wolbachia-negative* treatments (see Harcombe and Hoffmann 2004). Tetracycline-treated lines were given two generations to recover from the effects of antibacterial treatment prior to use in experiments. Separate flies were reared for *Wolbachia-positive* treatments and were not treated with tetracycline. PCR-detection assays were carried out to confirm the presence or absence of *Wolbachia* using primers for a *Wolbachia* surface protein (primers were: (F) 5'-TGGTCCAATAAGTGATGAAGAAAC -3', (R) 5'- AAAAATTAAACGCTACTCCA -3').

Flies of both *Wolbachia* conditions were placed in separate cages and allowed to lay eggs on grape juice plates supplemented with yeast paste for up to 4 hours. Eggs were then collected and chemically dechorionated with a two-fold dilution of sodium hypochlorite (NaOCl) at 4.8g per 100g (Commercial bleach) to remove microorganisms including viruses from the surface, washed thoroughly with distilled water, and grown at 25°C in standard fly medium supplemented with yeast. Three further generations were allowed to pass to avoid the risks of any possible confounding maternal effects. Virgin females were then isolated and aged for 3-4 days.

3.2.2. Immune challenge

Escherichia coli (DH10B strain, Invitrogen) and *Micrococcus luteus* (LZB055 strain, Blades Biological Ltd, Edenbridge, UK) were cultured overnight in lysogeny broth (1% tryptone, 0.5% yeast extract, 1% NaCl, ddH_2O) and nutrient broth (0.5% peptone, 0.3% yeast extract, 0.5% NaCl, ddH_2O) respectively, until OD was *ca.* 0.5 at a wavelength of 600nm. Bacteria aliquots were then frozen at -80°C in 10% glycerol until required. On the day before the experiment bacteria aliquots were thawed and re-cultured overnight. Cells were then washed three times and re-suspended in isotonic Ringers solution, and gently vortexed before injection. For *Drosophila* C Virus, a tissue culture infective $dose_{50}$ of 1000 $TCID_{50}$ in 69nl was provided by Ben Longdon (see Longdon, Fabian et al. 2012). *Plodia interpunctella* granulosis virus stock was donated by Amy Pederson (University of Edinburgh) and preparations contained 6.9×10^7 viral particles per ml (see Saejeng, Siva-Jothy et al. 2010 for full details). *Beauveria bassiana* was supplied by Matthew Tinsley (University of Stirling). Fungal spore concentration was 2×10^8 spores ml^{-1}, and the formulation was agitated briefly by gentle vortex prior to injection (see Tinsley, Blanford et al. 2006 for further details).

Flies were anaesthetised with CO_2, injected ventrally in the upper abdomen with 69nl of the appropriate challenge using a Nanoject II (Drummond scientific, Bromall, PA, USA),

immediately placed in vials containing a sugar-agar medium, and kept at 18°C for either 6 or 30 hours (6 hours was chosen as a time when heightened responses to immune challenges have been previously shown (Lemaitre, Reichhart et al. 1997; Shia, Glittenberg et al. 2009), and 30 hours was selected to monitor gene expression well after the challenge). Challenges to *Wolbachia-negative* flies were: unwounded control (anaesthetised only), wounded control (isotonic Ringers solution for *Drosophila*; 182mM KCl, 46mM NaCl, 3mM $CaCl_2$, 10mM Tris-Cl), *E. coli* (a gram-negative bacterium), *M. luteus* (a gram-positive bacterium), *Drosophila* C Virus (a natural *Drosophila* RNA virus), *Plodia interpunctella* granulosis virus (a DNA virus of the Indian meal moth), *B. bassiana* (an arthropod-infecting fungus; delivered in oil), and an oil control (87·5% Shellsol T, 12·5% Ondina EL). Challenges to *Wolbachia-positive* flies were: unwounded control (anaesthetised only), wounded control (isotonic Ringers solution), and *Drosophila* C Virus. Animals were placed in liquid nitrogen after 6 or 30 hours post-injection, and shortly after homogenised in TRIzol reagent (Invitrogen) and stored at -80°C pending RNA isolation. A total of 25 individual flies were used for each of 22 treatments. The entire experiment was replicated 6 times. However following initial Illumina sequencing results, only replicates 3 and 4 were sequenced. Thus, a total of 1,100 flies were used in this study.

3.2.3. RNA isolation, reverse transcription and quantitative PCR

RNA was extracted from whole insects using a standard TRIzol extraction protocol (Invitrogen). Total RNA was treated with RNase-Free DNase I (Ambion) to remove genomic DNA contamination, and complementary DNA was synthesised using a Promega reverse transcription system with random hexamers. Reverse transcription conditions were: 22°C for 10 minutes, 42°C for 60 minutes, and 95°C for 5 minutes. Real-time quantitative PCR (qRT-PCR) was carried out to determine whether immune responses were heightened following challenge. qRT-PCR assays were performed on an Applied Biosystems StepOnePlus cycler with the non-specific nucleic acid stain SYBR Green I (Applied Biosystems) used to

preferentially bind to dsDNA. Primers for fly ribosomal protein 49 (rp49) were used as the qRT-PCR reference in all cycler runs (rp49 primer sequences were: (F) 5' - GACGCTTCAAGGGACAGTACTTG - 3', (R) 5' - AAACGCCGTTCTGCATGAG - 3'). Immune markers were selected based on expected up-regulation of important immune genes in response to specific classes of pathogen. The anti-fungal peptide Drosomycin (primer sequences: (F) 5' - TCCGTGAGAACCTTTTCCAATATG - 3', (R) 5' - CCAGGACCACCAGCATCAG - 3') is induced in response to exposure to fungal peptides or gram-positive bacteria (Lemaitre, Reichhart et al. 1997; Rutschmann, Jung et al. 2000; Michel, Reichhart et al. 2001; Brennan and Anderson 2004; Zhang and Zhu 2009), the antibacterial polypeptide Diptericin (primer sequences: (F) 5' - GCTGCGCAATCGCTTCTACT - 3', (R) 5' - TGGTGGAGTGGGCTTCATG - 3') is induced in response to gram-negative bacteria (Lemaitre, Reichhart et al. 1997; Rutschmann, Jung et al. 2000; Brennan and Anderson 2004), and Vir-1 (primer sequences: (F) 5' - GATCCCAATTTTCCCATCAA - 3', (R) 5' - GATTACAGCTGGGTGCACAA - 3'), a virus-induced target gene of the Jak/STAT host antiviral signalling pathway (Dostert, Jouanguy et al. 2005; Ip 2005) is shown to be up-regulated in response to DCV (Dostert, Jouanguy et al. 2005; Hedges and Johnson 2008). All cycler runs included serial dilutions of fly cDNA to calculate primer efficiency, controls for the presence of genomic DNA, and a melting curve analysis for product identification. Each reaction contained: 8µl Fast SYBR® Green Master Mix (Applied Biosystems), 10µl ddH_2O, 1µl primer mix (10mM concentration), and 1µl template cDNA. qRT-PCR conditions were: holding stage: 95ºC for 10 minutes; cycling stage: 40 cycles of 95ºC for 15 seconds, 60ºC for 60 seconds; and melt curve stage: 95ºC for 15 seconds, 60ºC for 60 seconds with temperature increment +0.3ºC for 15 seconds to 95ºC. The qRT-PCR assay included 3 technical replicates and two biological replicates. A further qRT-PCR assay was carried out to determine if immune responses had occurred at the time points chosen to harvest flies post-challenge. As such, flies were exposed to either *E. coli* or *M. luteus* and the expression levels of two classical

immune marker genes (Diptericin and Drosomycin) were quantified across a time-course. Fly preparation, immune challenge, RNA isolation and qRT-PCR conditions were as described previously.

3.2.4. PCR and Illumina sequencing

Unique molecular ID (MID) tags were included on the 5' end of all forward primers to identify treatments in multiplexed sequencing reactions. Primers were designed to amplify a fragment of *D. melanogaster Dscam* spanning from exon 3 to exon 7 (primer sequences were: (F) 5' - ACGTCCATGTGCGAGCCGT - 3', (R) 5' - GTTCCTTCGATGAACTTGTACCAT - 3'). Thus, PCR amplicons comprised variable exons 4 and 6, and the conserved exon 5. PCR conditions were: 98°C for 30 seconds, then 35 cycles of 98°C for 30 seconds, 55°C for 10 seconds, and 72°C for 60 seconds, and finally 72°C for 10 minutes. PCR products were quantified using the Qubit® 2.0 Fluorometer (Invitrogen), and all MID-tagged samples were mixed evenly based on quantification. The final mixed product was concentrated using ethanol precipitation, and sequenced in an Illumina Genome Analyzer II (service provided by the GenePool Sequencing Facility, University of Edinburgh).

3.2.5. Cloning and sequencing

Cloning and sequencing was carried out on two different treatments to compare with Illumina sequencing data. PCR products were cloned using a TOPO TA cloning kit with pCR®4-TOPO® vector and transformed in chemically competent *E. coli*. Around 100 clones per treatment were sequenced in a 96-capillary ABI 3730*xl* DNA Analyzer (service provided by the GenePool Sequencing Facility, University of Edinburgh). Sequences were aligned using BioEdit version 7.0.9.0 and identified by cross-referencing with the known *D. melanogaster* genome sequence available from the Ensembl genome browser (http://www.ensembl.org/index.html).

3.2.6. Data analysis

Raw data handling

Sequencing data from the Illumina GAII was received as FASTQ files (containing multiple short-read sequences with quality scoring information), and was converted to FASTA files with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Pre-processing of reads was carried out to produce better alignment results. A Perl Script was used to cut bad quality ends (ConDeTri v2.0 (Smeds and Kunstner 2011); content dependent read trimming software for Illumina/Solexa sequencing data). This involved the removal of non-ATGC characters and the trimming of poor quality sequence by setting stringency parameters. High and low quality thresholds were set at 25 and 20, respectively (Sanger Scale; $Q_{sanger} = -10 \log_{10} p$, where p is the probability of the corresponding base call being incorrect), and the fraction of the read which was to exceed high quality was set at 80%. Minimum read length was set at 50bp and reads without a valid MID-tag were discarded. Reads were identified by cross-referencing with the *D. melanogaster* genome (http://www.ensembl.org/index.html).

Statistical analyses

Quantification of immune marker expression in qRT-PCR assays: To approximate the change in expression of immune markers from qRT-PCR assays, mean cycle threshold (Ct) values for each treatment were firstly calculated from three technical replicates, and then a delta Ct (ΔCt) value was calculated for each treatment as the difference between the cycle thresholds (Ct) of the immune marker genes (Diptericin, Drosomycin or Vir-1) and the endogenous control gene (rp49). Relative difference between immune markers and rp49 was expressed as $2^{-\Delta Ct}$ and *mean relative expression* ± SE was calculated from two biological replicates. Average primer efficiencies for Diptericin, Drosomycin, Vir-1 and rp49 were 101%, 99%, 91% and 100%, respectively. To approximate the fold change of immune marker expression over a time-course, the delta delta Ct (ΔΔCt) method was used (Livak and Schmittgen 2001). Mean Ct of each treatment was firstly calculated from three technical

replicates. Delta Ct (ΔCt) was then calculated as the difference between mean Ct values of marker genes (here, Diptericin and Drosomycin) and rp49 for each time point, and fold change of expression compared to unchallenged samples was calculated as $2^{-\Delta\Delta Ct}$. *Mean fold change* ± SE was calculated from three biological replicates. Average primer efficiencies for Diptericin, Drosomycin, and rp49 were 109%, 104% and 103%, respectively.

Analysis of constitutive exon expression: I tested for a departure from even expression using chi-square (χ^2) tests on the mean raw counts of exons 4 and 6 of unchallenged *Drosophila* to determine whether the frequency distribution of each exon variant was consistent with a random expression model (i.e. all exon variants are equally likely to be expressed; expectation calculated as 1/n variants at a given exon cluster). Exon 4-6 combinations in unchallenged *Drosophila* were tested for even expression using Fisher's Exact Tests on raw count data for each unwounded treatment due to low counts for some combinations. Generalised linear mixed models using MCMCglmm (Hadfield 2009) were also implemented to reveal non-random constitutive expression by providing estimates of the proportion of variance explained by exons 4 and 6. See below for description of MCMCglmm package. All analyses were handled in R (R Development Core Team 2011).

Analysis of infection-responsive Dscam splicing in Drosophila: I used three approaches to detect infection-responsive splicing in *Drosophila*. Firstly, indices of diversity were calculated from count data using the Simpson's Index of Diversity (see Chapter 2, and also Simpson 1949), and I fitted a GLM to the diversity indices to test for effects of time and pathogen, or an interaction between both.

I then used two complementary methods to examine Dscam expression in response to immune challenge. Firstly, to infer the effect of immune challenge on Dscam expression patterns, I used the MCMCglmm (Markov chain Monte Carlo methods for generalised linear

mixed models) R/Bioconductor package v2.16 (Hadfield 2009). MCMCglmm combines Markov chain Monte Carlo (MCMC) sampling methods with generalised linear mixed models (GLMMs). Generalised linear models (GLM's) extend simple regression to allow for the response variable having a non-Gaussian distribution when modelling of a relationship between experimental variables (Guisan, Edwards et al. 2002), while generalised linear *mixed* models (GLMM's) extend this again by allowing explanatory variables to contain random as well as fixed effects (Breslow and Clayton 1993). Markov chain Monte Carlo methods are algorithms which sample from different probability distributions and follow a Markov Chain; a random process considered to be memory-less because as it undergoes transitions from one state to another during simulations, its next state depends only on its current state, and not of the prior state (Hastings 1970). As such, MCMCglmm fits GLMM's using MCMC methods to find sources of variation, then estimate the proportion of variance explained by each variable (e.g. estimate the proportion of total variance explained by random effects).

I used MCMCglmm to infer the effect of wounding or parasite-exposure on Dscam expression patterns. Specifically, I tested for changes in Dscam expression between (1) *Wolbachia*-cleared unwounded and *Wolbachia*-cleared Ringers-wounded treatments (i.e. to investigate any tissue-damage effects on Dscam expression) and whether there was any interaction with time (i.e. at 6 or 30 hours), (2) different pathogen treatments compared to sterile-wounded treatments (i.e. to investigate any effects of pathogen-exposure) and whether there was any interaction with time, and (3) Ringers and DCV treatments in both *Wolbachia*-positive and *Wolbachia*-cleared flies to test for differences in Dscam expression in response to DCV between the *Wolbachia* infection states. To test for effects of wounding, I fitted the following model to exon read counts: reads ~ time + wounding + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: time + exon 4: wounding + exon 6: time + exon 6: wounding + exon 4: time: wounding + exon 6: time: wounding + exon 4-6: time + exon 4-6: wounding +

exon 4-6: time: wounding. Time and wounding were treated as fixed effects and pathogen-exposed and *Wolbachia*-positive treatments were excluded from this analysis. To test for effects of pathogen-exposure, I fitted the model: reads ~ time + pathogen + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: time + exon 4: pathogen + exon 6: time + exon 6: pathogen + exon 4: time: pathogen + exon 6: time: pathogen + exon 4-6: time + exon 4-6: pathogen + exon 4-6: time: pathogen. Time and pathogen were treated as fixed effects and unwounded and *Wolbachia*-positive treatments were excluded from this analysis. To test for effects of DCV-exposure in treatments with and without *Wolbachia*-infection, I fitted the model: reads ~ DCV + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: *Wolbachia* + exon 4: DCV + exon 6: *Wolbachia* + exon 6: DCV + exon 4: *Wolbachia*: DCV + exon 6: *Wolbachia*: DCV + exon 4-6: *Wolbachia* + exon 4-6: DCV + exon 4-6: *Wolbachia*: DCV. The pathogen was treated as a fixed effect and time and unwounded treatments were excluded from this analysis. Relative expression was normalised as cube-root of percentages (expressed as reads per million) to ensure normally distributed residuals (see appendix figure S4 for Q-Q plot of distribution before and after transformation). Two runs of 1,000,000 MCMC iterations were completed for each model. Each run was sampled every 100 steps and the first 10% of iterations was discarded as burn-in.

I also used the R/Bioconductor package DESeq (v1.8.1) to analyse the discrete count data assigned to *Dscam* exon variants across different libraries. Count data are skewed and hence not well approximated by continuous distributions, especially for small samples and in lower count ranges (Oshlack, Robinson et al. 2010). Accordingly, DESeq tests for differential expression and accounts for biological variability by use of the negative binomial distribution and a shrinkage estimator for the distribution's variance (Anders and Huber 2010). DESeq also accounts for sampling variance (the variance in numbers of reads that persists even if everything is exactly equal), technical noise (introduced by sample preparation and sequencing) and biological noise (from differences between samples)

(Anders and Huber 2010). DESeq is used to analyze count data associated with different genes. In this study, numbers of reads were assigned to Dscam transcripts for each treatment and gene variants were treated as equivalent to genes for the purposes of DESeq analysis. Within-library normalisation was then carried out to remove technical effects. Normalisation by sequencing-depth usually adjusts by the total number of reads in the library (i.e. library size-scaling) which is expected to account for differences in sequencing-depth between samples, but does not account for either RNA composition effects or that a small number of highly expressed variants can consume a significant amount of the total sequence and may distort the ratio of total reads (Anders and Huber 2010; Oshlack, Robinson et al. 2010). As such, scaling factors were estimated for each library to allow more confident quantification of expression levels of each variant relative to other variants in the sample. Here, the sequencing-depth is estimated by the count of the *Dscam* exon variant with the median count ratio across all variants. Counts are multiplied by scaling factors and scaled data are used for detecting differential expression.

Here, using DESeq, I tested for differential expression at 6 and 30 hours between (1) unchallenged and sterile-wounded flies with and without *Wolbachia* infection to test for an effect of wounding on exon 4 and 6 variant-expression, (2) sterile-wounded and septic-wounded treatments to test for an effect of pathogen-exposure on the expression of exon 4 and 6 variants, (3) sterile-wounded and DCV-exposed flies either harbouring or cleared of *Wolbachia* bacteria to test for a difference in Dscam expression between the infection states, (4) unchallenged flies and septic-wounded flies to examine the difference between significantly different exposures, and (5) comparisons 1-4 above but testing the expression of exon 4-6 combinations. DESeq corrects for multiple testing using Benjamini-Hochberg's FDR (Benjamini and Hochberg 1995), and here FDR was controlled at 5% ($q = 0.05$). All analyses were handled in R (R Development Core Team 2011).

Contingency tests: Finally, the mean proportions of expression of exons 4 and 6 between the Illumina sequencing data and those of the cloning sequencing assay were analysed using the Fisher's Exact Test. The aim here was to test for an association between the data of two independent sequencing methods on the same RNA samples. All tests were handled in R (R Development Core Team 2011).

3.3. Results

3.3.1. qRT-PCR was used to confirm immune responses had occurred in challenged flies

Firstly, to confirm whether immune responses occurred in the flies used in this study, I carried out qRT-PCR assays to detect changes in expression levels of common immune markers associated with exposure to specific immune elicitors. Figure 2 shows the mean expression of immune marker genes relative to an endogenous control gene.

I found levels of Diptericin highest in *E.coli*-exposed treatments, with notable increases also seen post-exposure to *M. luteus* at 6 hours post-injection (*pi*) and to PiGV at 30 hours *pi* (Fig. 2). No other strong effects on the expression of Diptericin were found. Thus, a relatively strong Diptericin response appeared to occur in flies exposed to *E. coli*, with smaller but notable effects on expression also seen post-exposure to *M. luteus* and PiGV. Rises in relative expression levels of Drosomycin were found in treatments exposed to the gram-positive bacterium *M. luteus* at 6 hours *pi* with a further increase at 30 hours *pi* (Fig. 2). Moreover, relatively strong effects of *E.coli*-exposure were also seen on Drosomycin expression (Fig. 2). Thus, flies exposed to *M. luteus* and also *E. coli* appeared to induce strong Drosomycin responses against the challenges. Exposure of flies to *B. bassiana* did not appear to up-regulate Drosomycin. This could have been due to the injection of the fungus in an oil suspension which may have suppressed any immune response. Consequently, these

treatments and the complementary oil controls were not included in subsequent statistical analyses. Vir-1 levels have been shown to increase in the presence of virus in *Drosophila* (Dostert, Jouanguy et al. 2005). Here, no convincing effects of DCV-exposure could be seen on the relative expression levels of Vir-1 (Fig. 2). PiGV also had little effect on the expression of vir-1 compared to non-viral immune challenges, but the virus had an unexpected effect on Drosomycin expression. This suggested some form of contamination may have been present. As a consequence PiGV treatments were also disregarded in subsequent statistical analyses. Thus, no convincing effects of viral-exposure were found in this study following qRT-PCR.

Notably, sterile-wounding and DCV-exposure in *Wolbachia*-positive flies resulted in higher expression of *all* markers than in corresponding *Wolbachia*-cleared treatments. This suggested that the immune responses were stronger in *Wolbachia*-infected flies. Accordingly, this observation encouraged further analyses on DCV-challenged treatments of both *Wolbachia* infection-states.

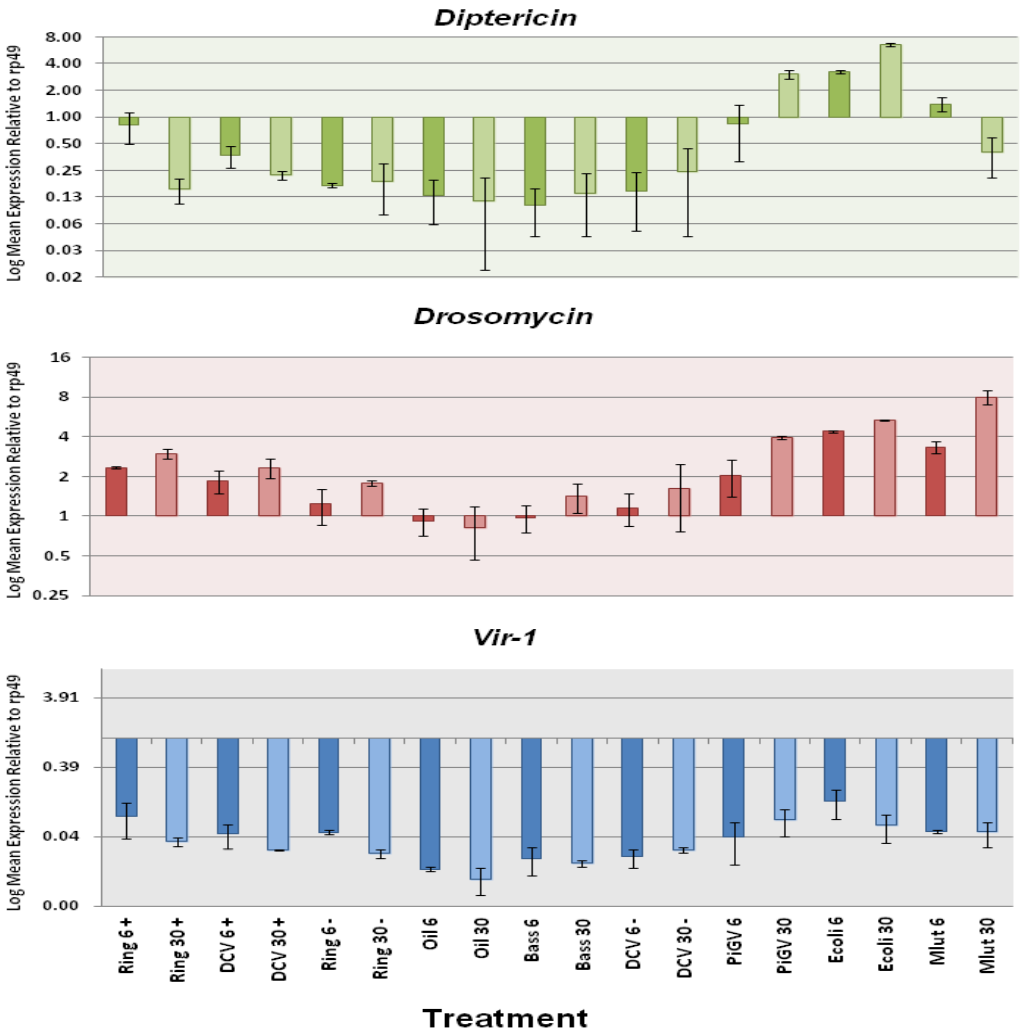


Figure 2: Mean expression levels of immune markers relative to those of ribosomal protein 49 (rp49). Levels of Diptericin (top) were highest in *E. coli* treatments with notable increases in expression also seen in response to *M. luteus* at 6 hours, and to PiGV at 30 hours. Drosomycin (middle) showed increases in expression levels in response to challenge at 6 hours post-exposure in most treatments and in all cases expression levels had further increased by 30 hours (except in oil controls). Up-regulation of Drosomycin followed exposure to both bacteria-types as expected (Lemaitre, Reichhart et al. 1997), however exposure to *B. bassiana* had little effect. Expression of Vir-1 (bottom) showed minimal up-regulation across treatments. Notably, sterile-wounding and DCV-exposure in *Wolbachia*-positive flies resulted in higher expression of all three markers than in *Wolbachia*-cleared flies. All values are expressed on a log base 2 scale. Error bars represent SE calculated from two biological replicates. **Key**: +: *Wolbachia*-positive, -: *Wolbachia*-cleared, Ring: Ringers, DCV: *Drosophila* C Virus, Oil: oil control, Bass: *B. bassiana*, PiGV: *Plodia interpunctella* granulosis virus, Ecoli: *Escherichia coli*, Mlut: *Micrococcus luteus*, 6: 6 hours, 30: 30 hours.

3.3.2. Illumina sequencing assay output

Data received from the sequencing facility had 13,326,557 read-pairs. After quality trimming (see Methods), I had 11,480,363 reads which were still available in pairs of reads at least 50nt long. Sequences without a valid MID-tag or affected by mis-priming or template switching were then discarded and a total of 8,983,816 reads remained and were attributable to exons 4 and 6, and combinations thereof. The mean number of high quality reads from each PCR primer-pair was 187,162 (sequencing-depth ranged from 59,373 to 295,885).

3.3.3. Detecting differences in constitutive exon variant expression, detecting novel exon variants, and identifying unexpressed exons

I used Pearson's chi-square and MCMCglmm analyses to test for uneven expression of exons 4 and 6 in different tissues of unchallenged flies. I found significant differences in the relative expression between variants of both exons 4 (χ^2 = 29147.1, DF = 11, p = <0.0001) and 6 (χ^2 = 39209.5, DF = 47, p = <0.0001) across all unwounded treatments (Figure 3, see left columns). Using MCMCglmm, I found significant effects of exon 4- and exon 6-expression (Table 1), indicating non-random expression of both exons. Thus, variants of exons 4 and 6 appear to be expressed in a non-random fashion in *Drosophila*.

Exon 4.2 was by far the most abundant exon 4 variant expressed (Fig. 3), and exon 4.9 was the least abundant exon 4 variant. Furthermore, exons 4.4, 4.10 and 4.11 were under-represented compared to expected levels (in a random distribution, i.e. 1/number of alternative variants at a given exon) in unwounded flies (Fig. 3). For exon 6, variant 6.25 was the most abundant exon 6 variant (Fig. 3). Several exon 6 variants were over-represented compared to expected levels in unwounded flies, while others were under-represented (Fig. 3).

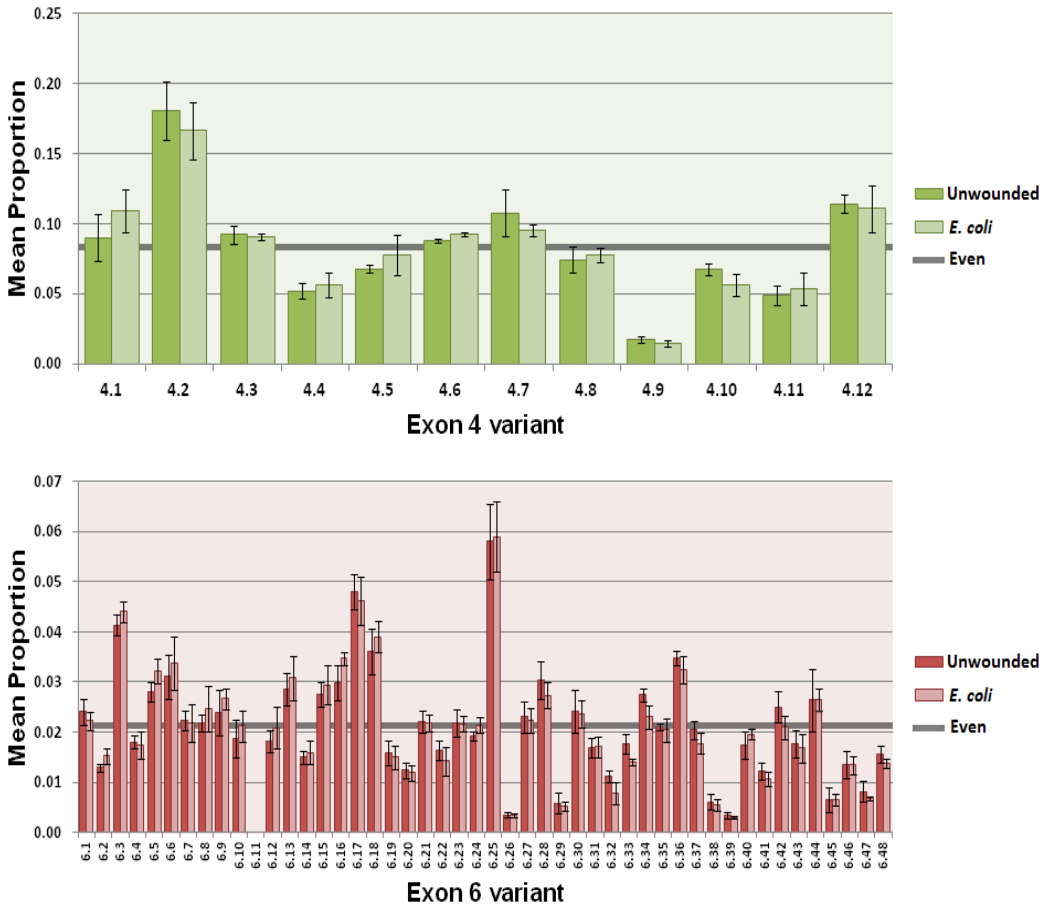


Figure 3: Bar chart representing relative expression between different variants of exons 4 and 6 in unchallenged *D. melanogaster* compared to expression in flies challenged with *E. coli*. (**A**) Exon 4 variant expression (top) in unwounded *Wolbachia*-cleared treatments (see left bars) shows significant differences between the relative expression of individual variants (χ^2 = 29147.1, DF = 11, p = <0.0001). Exon 4.2 is the most abundant variant in unchallenged adult flies, while exon 4.9 is the least abundant. Similarly, exon 6 gene variant expression showed significant non-random expression (χ^2 = 39209.5, DF = 47, p = <0.0001). Exon 6.25 was the most abundantly expressed exon 6 variant, while exon 6.39 was the least abundant. Exon 6.11 was entirely absent. Expected proportion is shown as a horizontal line and based on a random expression of all variants equally (i.e. even expression = 1/number of variants at a given exon cluster). (**B**) Gene variant expression was also compared between *E. coli*-challenged flies (right bars) and unwounded flies (left bars). No statistically significant differences in the expression of any exon 4 variant (top) (DESeq, q-values ≥ 0.44), or any exon 6 variant (bottom) (q-values ≥ 0.52) was found between the treatments. All bars represent expression at 6 hours. Error bars for unwounded treatments are standard deviation (n=4), and error bars for *E. coli* treatments are standard deviation (n=4).

Notably, exon-skipping was detected in unchallenged flies. Exon 6-skipping was detected at a frequency of $1.3x10^{-3}$. Exon 4-skipping was not expected as the 5' primer nested in constitutive exon 3 overlapped variable exon 4 by one base. Nevertheless, a low frequency of exon 4-skipping was detected.

To investigate whether our sequencing assay could detect novel exon variants, sequences in the data set were compared to those of the reference genome for *D. melanogaster*. Alternative alleles not represented in the reference genome were detected. Two allelic variations of exon 6.12 were found, containing one Single Nucleotide Polymorphism (SNP), and two allelic variations of exon 6.39 were found, which also differed from the reference genome at one base. In contrast, no allelic variations of exon 4 were detected. Overall, all possible exon 4 variants were expressed in all treatments in this study, while all exon 6 variants were also detected with the exception of exon 6.11 which was entirely absent.

3.3.4. Detecting constitutive combinatorial splicing in unwounded Drosophila and determining whether splicing is random or non-random

I next tested whether the assay could detect combinatorial splicing in unchallenged flies and determine whether splicing is random or non-random in the absence of wounding or parasites. Figure 4 shows a heat map of relative expression levels of all possible exon 4-6 combinations and visual inspection suggested that the splicing of these two exons is close to random (i.e. conditional on the frequency of each exon 4 and 6 variant independently). For example, as would be expected from random alternative splicing, exon-exon combinations containing variants 4.2 and 6.25 were the most highly abundant as these were the most abundant variants for each exon cluster independently (expected frequency was 0.184 x 0.058 = 0.011, actual frequency was 0.011), and represented just over 1% of all reads (Fig. 4). The least abundant exon 4 and 6 variants, 4.9 and 6.39 respectively, were also the least abundant combination as expected in a random splicing model and represented less than

0.0001% of all reads (Fig. 4). Additionally, the frequency of expression of combination 4.1-6.1 was expected based on independent frequencies of exons 4.1 and 6.1 (expected = 0.09 x 0.025 = 0.002, actual = 0.002). However, some combinations appeared to show non-random splicing. Based on the independent frequencies of exons 4.4 and 6.31 for example, the expected frequency of this combination would be around 0.0009 (0.050 x 0.017), but the actual frequency was 0.0014, or nearly 60% higher than expected (Fig. 4). Additionally, the independent frequencies of 4.10 and 6.9 would suggest the two gene variants would combine at a frequency of 0.0016 (0.061 x 0.026), however the actual frequency was 0.0010, or around 60% lower than expected (Fig. 4). Following analysis with MCMCglmm, I found a small exon 4-6 combination effect (Table 1), indicating that a small number of combinations were non-randomly expressed. Thus, exon 4-6 combinatorial splicing appeared to be close to random in unwounded flies, however some combinations appeared to be expressed in a non-random fashion.

Overall, all possible exon 4-6 combinations were detected with the exception of those necessitating exon 6.11, resulting in a total of 564 exon 4-6 combinations expressed. Thus, high Dscam transcript diversity was present in unwounded flies which demonstrated a non-restricted expression for both studied exons under the background of the whole insect.

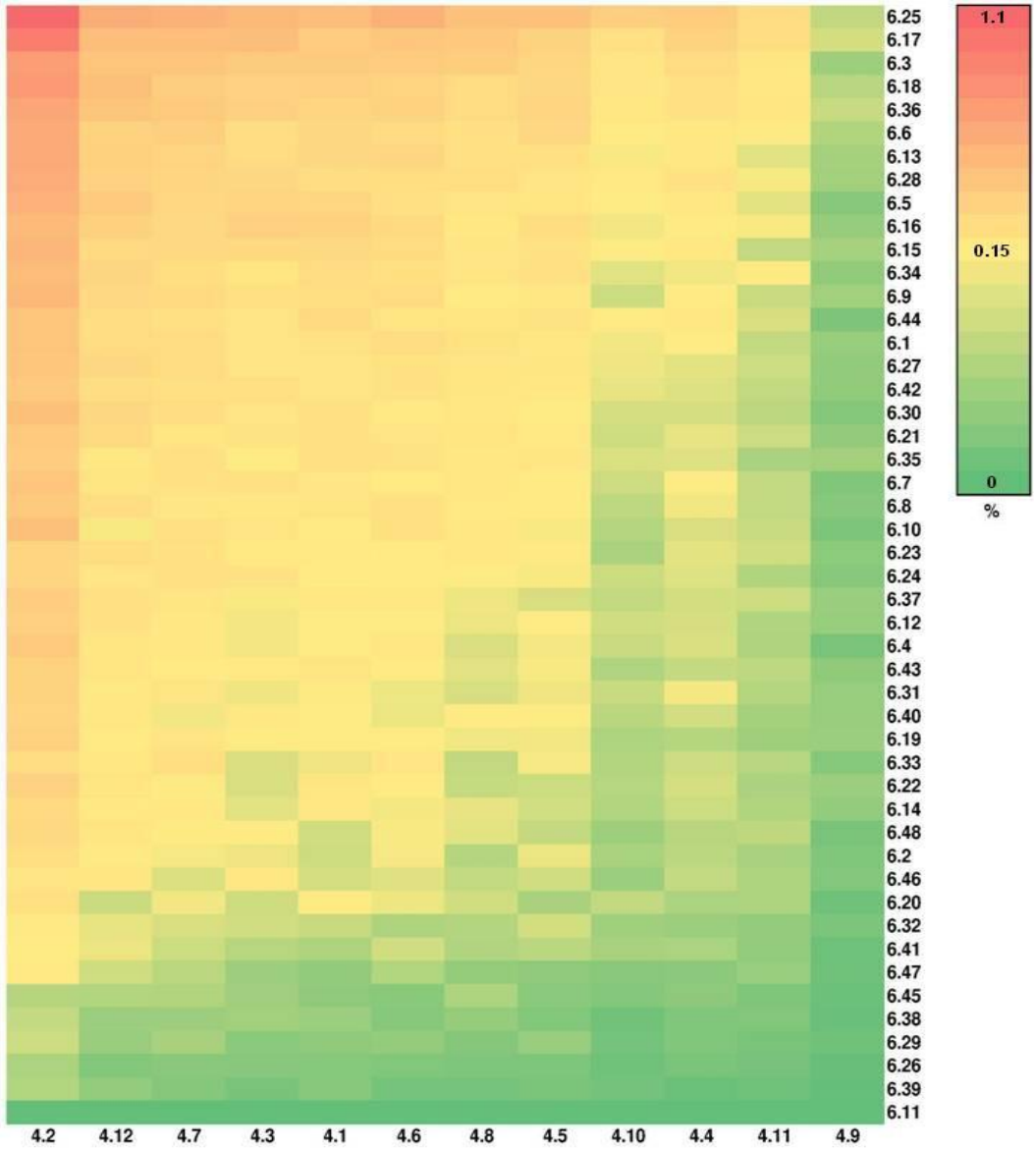


Figure 4: Heat map representation of the relative expression of all possible exon 4-6 combinations in unwounded *D. melanogaster*. Combinatorial splicing appeared to be close to random. The most abundant combination was 4.2-6.25, which represented just over 1% of all sequencing reads obtained, and was expressed at an expected frequency based on independent frequencies of variants 4.2 and 6.25. Some combinations appeared to show non-random association however. For example, combination 4.4-6.31 was expressed around 60% higher than expected, and combination 4.10-6.9 was represented around 60% lower than expected. Indeed, Fisher's Exact Tests carried out on the raw data counts of each exon combination found deviation from even expression in all unwounded treatments in the study (all, $p = < 0.0001$). Each cell represents the proportion of an exon 4-6 combination, calculated as the mean of biological replicates of unwounded treatments (n=10).

3.3.5. Testing for differential Dscam expression in response to a wide array of immune elicitors in flies with or without Wolbachia infection

Firstly, in line with the analysis carried out in Chapter 2, indices of diversity were calculated for exon 4, exon 6, and exon 4-6 combinations for each treatment in the study. As expected, all indices were high due to the high diversity of Dscam isoform expression (see Table S4). I fitted a GLM to expression diversity data as: Simpson's Index of Diversity (1-D) ~ time + pathogen + time*pathogen. For exon 4 expression-diversity, I found no significant effects of time ($F_{1,33} = 2.58$, $p = 0.122$), pathogen ($F_{10,33} = 1.91$, $p = 0.098$), or an interaction between both ($F_{2,32} = 0.85$, $p = 0.588$). Equally, no significant effects were found for exon 6 (time: $F_{1,33} = 1.92$, $p = 0.180$; pathogen: $F_{10,33} = 0.66$, $p = 0.748$; interaction: $F_{2,32} = 1.25$, $p = 0.314$), or exon 4-6 combinations (time: $F_{1,33} = 1.20$, $p = 0.285$; pathogen: $F_{10,33} = 1.68$, $p = 0.148$; interaction: $F_{2,32} = 0.85$, $p = 0.589$). Thus, expression diversity was not significantly affected by pathogen-exposure at either 6 or 30 hours.

Using MCMCglmm, I fitted models to the count data to test for an interaction between Dscam expression and experimental conditions. I firstly tested for effects of wounding. Table 1 outlines the analysis output which represents the variance structure (i.e. the proportions of variance explained by each of the variables in the model). The model mixed well (i.e. the model explored parameter space efficiently and effectively) and the average effective sample size was around 9000 (see appendix, figure S5). I found that a considerable proportion of the total variance was explained by exon 4 and by exon 6, with a relatively smaller proportion explained by exon 4-6 combination expression (Table 1). Otherwise, contributions to total variance were very low for all terms in the model and lower credibility bounds were very close to zero. Thus, non-random expression of exons 4 and 6 was detected but no significant effects of wounding were evident.

Table 1: Variance structure summarising MCMCglmm analysis of a model testing for effects of wounding and time on Dscam expression. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Primer-pair	0.031	0.015	0.012	[0, 0.10]	0.1
Exon 4	**11.919**	**10.230**	**7.758**	**[3.48, 24.10]**	**56.9**
Exon 6	**6.053**	**5.871**	**5.180**	**[3.81, 8.74]**	**28.9**
Exon 4-6	**0.129**	**0.129**	**0.129**	**[0.07, 0.20]**	**0.6**
Exon 4: Time	0.050	0.036	0.001	[0, 0.14]	0.2
Exon 4: Wounding	0.077	0.059	0.001	[0, 0.20]	0.4
Exon 6: Time	0.013	0.009	0.0002	[0, 0.04]	0.1
Exon 6: Wounding	0.012	0.006	0.0002	[0, 0.04]	0.1
Exon 4: Time: Wounding	0.052	0.046	0.0280	[0, 0.12]	0.2
Exon 6: Time: Wounding	0.032	0.030	0.0002	[0, 0.07]	0.2
Exon 4-6: Wounding	0.013	0.007	0.0001	[0, 0.05]	0.1
Exon 4-6: Time	0.016	0.009	0.0003	[0, 0.06]	0.1
Exon 4-6: Time: Wounding	0.012	0.005	0.00008	[0, 0.04]	0.1
Residuals	2.559	2.558	2.551	[2.44, 2.68]	12.2

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

I next tested for effects of pathogen. Table 2 outlines the analysis output representing the variance structure for the model. The model mixed well and the average effective sample size was around 9000 (see appendix, figure S6). In addition to the variance explained by the non-random expression of exons 4 and 6, I found a very small exon 4-by-time-by-pathogen effect, which contributed 0.5% to the total variance explained by random effects in the model and the lower credibility bound did not reach zero (Table 2). Otherwise, I found no discernible effects of pathogen-exposure on the expression of exons 4 and 6, or their interaction with any terms in the model.

Table 2: Variance structure summarising MCMCglmm analysis of a model testing for effects of pathogen-exposure and time on Dscam expression. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Primer-pair	0.014	0.012	0.007	[0, 0.03]	0.1
Ex4	**15.857**	**13.928**	**11.398**	**[5.59, 31.65]**	**64.5**
Ex6	**6.168**	**6.002**	**5.424**	**[3.80, 8.85]**	**25.1**
Exon 4-6	**0.186**	**0.186**	**0.187**	**[0.15, 0.23]**	**0.8**
Ex4: Time	0.012	0.005	0.0005	[0, 0.05]	0.0
Ex4: Pathogen	0.034	0.028	0.0004	[0, 0.09]	0.1
Ex6: Time	0.002	0.001	0.00002	[0, 0.01]	0.0
Ex6: Pathogen	0.009	0.008	0.00007	[0, 0.02]	0.0
Ex4: Time: Pathogen	**0.117**	**0.115**	**0.113**	**[0.06, 0.17]**	**0.5**
Ex6: Time: Pathogen	0.008	0.006	0.0001	[0, 0.02]	0.0
Exon 4-6: Pathogen	0.005	0.002	0.00004	[0, 0.02]	0.0
Exon 4-6: Time	0.009	0.005	0.00008	[0, 0.03]	0.0
Exon 4-6: Time: Pathogen	0.003	0.002	0.00005	[0, 0.01]	0.0
Residuals	2.161	2.161	2.159	[2.10, 2.23]	8.8

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

Lastly, I tested for effects of DCV-exposure in treatments with and without *Wolbachia*-infection. Table 3 outlines the analysis output representing the variance structure for the model. The model mixed well and the average effective sample size was around 5000 (see appendix, figure S7). Here, a very small exon 4-by-*Wolbachia*-by-DCV effect was found which explained around 0.2% of the total variance. In addition, a similarly sized proportion of the total variance was unexpectedly explained by primer-pair (Table 3). The lower credibility bounds of these effects did not reach zero but were very low. Otherwise, there was no evidence of an effect of DCV-exposure on Dscam expression.

Table 3: Variance structure summarising MCMCglmm analysis of a model testing for effects of DCV-exposure on Dscam expression in *Wolbachia*-positive and *Wolbachia*-cleared treatments. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Primer-pair	**0.007**	**0.006**	**0.005**	**[0.002, 0.014]**	**0.2**
Ex4	**3.166**	**2.827**	**2.349**	**[1.134, 6.039]**	**77.4**
Ex6	**0.652**	**0.632**	**0.610**	**[0.400, 0.943]**	**16.0**
Exon 4-6	0.002	0.002	0.003	[0, 0.005]	0.1
Ex4: *Wolbachia*[3]	0.004	0.003	0.0001	[0, 0.012]	0.1
Ex4: DCV[3]	0.003	0.002	0.00002	[0, 0.010]	0.1
Ex6: *Wolbachia*	0.0003	0.0002	0.00001	[0, 0.001]	0.0
Ex6: DCV	0.0009	0.0008	0.00002	[0, 0.002]	0.0
Ex4: *Wolbachia*: DCV	**0.007**	**0.007**	**0.006**	**[0.002, 0.013]**	**0.2**
Ex6: *Wolbachia*: DCV	0.0003	0.0001	0.000002	[0, 0.001]	0.0
Exon 4-6: DCV	0.001	0.0008	0.00002	[0, 0.004]	0.0
Exon 4-6: *Wolbachia*	0.0006	0.0003	0.00001	[0, 0.002]	0.0
Exon 4-6: *Wolbachia*: DCV	0.0007	0.0004	0.00001	[0, 0.002]	0.0
Residuals	0.244	0.244	0.244	[0.237, 0.251]	6.0

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.
[3] *Wolbachia* (yes/ no), DCV (yes/ no).

I also used DESeq and MCMCglmm to test the number of reads for each *Dscam* exon variant for being differential between treatments. Figure 5 shows heat maps representative of exon 4- and exon 6-expression for all treatments in the study. Strong patterns were seen for exon variants across all treatments for both exons 4 and 6 (Fig. 5; columns), however visual inspection suggested that no patterns were obvious between different pathogens or pathogen types (Fig. 5; rows).

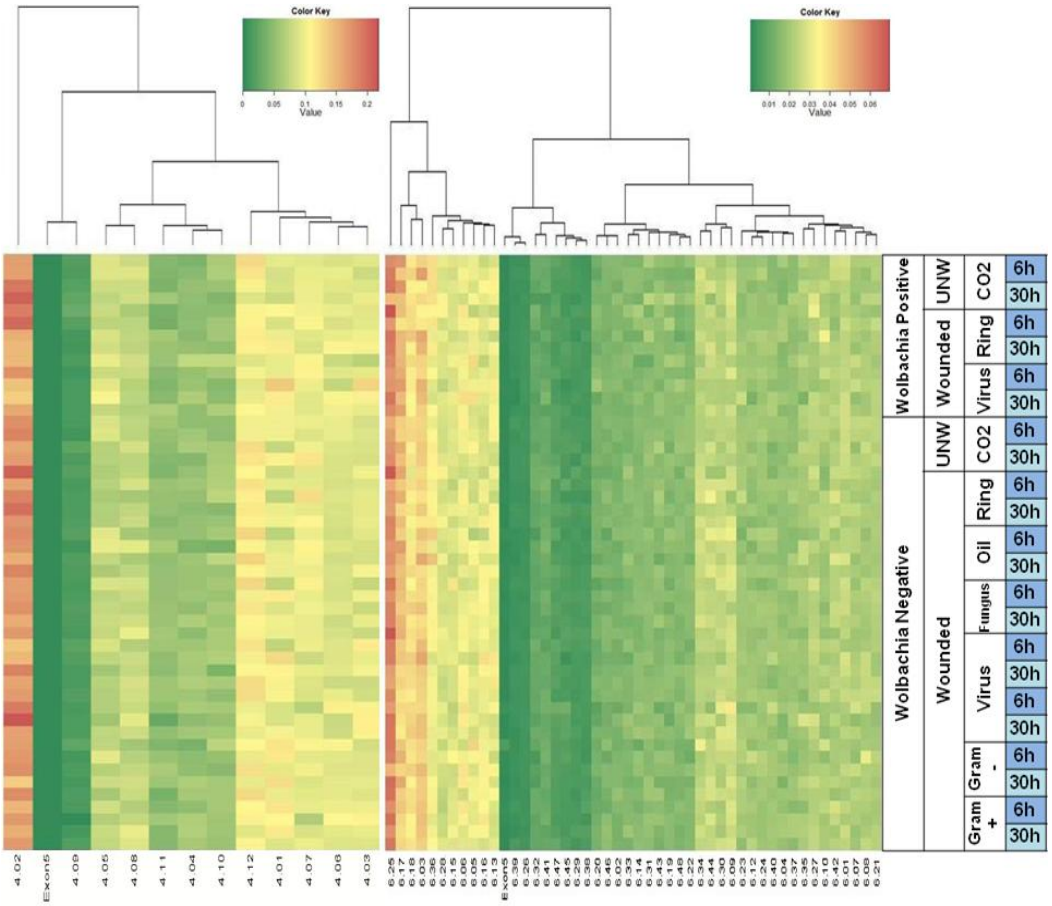


Figure 5: Heat map representation of relative expression of exon 4 (left block) and exon 6 (right block) variants for all treatments in this study. Patterns of expression for individual variants of both exons 4 and 6 across treatments were evident. For example, exon 4.2 was clearly the most abundantly expressed exon 4 variant across all treatments, while exon 4.9 was the least abundant exon 4 variant across all treatments. For exon 6, exon 6.25 was expressed higher than any other across all treatments, with 6.39 showing the lowest relative expression levels across treatments. No discernible patterns were obvious between treatments. Exon 6.11 was not detected and hence not included. The inclusion of exon 5 indicates relative abundance of exon-skipping for each exon cluster. Each time point (6h, 30h) includes 2 biological replicates, and each cell represents the proportional expression of a given exon variant in a single sample. Dendrograms result from hierarchical clustering of exon variant-expression pattern using Euclidean distance measures. **Key**: *Wolbachia* Positive: flies harbouring *Wolbachia*, *Wolbachia* Negative: flies cleared of *Wolbachia*, UNW: Unwounded, CO2: carbon dioxide anaesthetised, Ring: Ringers solution, Oil: oil control, Fungus: *B. bassiana*, Virus: DCV and PiGV, Gram -: *E. coli*, Gram +: *M. luteus*, 6h: 6 hour time point, 30h: 30 hour time point.

Using DESeq, I found no evidence of differential expression between any pairwise comparisons, except that exons 4.2 and 4.9 were significantly lower in *Wolbachia*-infected flies at 30 hours following sterile-wounding (both q-values = 0.04), and exon 4.9 was significantly lower in *Wolbachia*-cleared flies following *M. luteus*-exposure at 6 hours (q = 0.02). Consistent with DESeq, analysis with MCMCglmm detected a significant exon 4.9-effect at 6 hours following *M. luteus*-exposure in *Wolbachia*-cleared flies (p = 0.0001). I also compared the relative expression of exons 4 and 6 between unwounded treatments and septic-wounded treatments exposed to *E. coli* (Fig. 3, see both columns). *E. coli* was the immune elicitor which resulted in the most significant up-regulation of its immune marker as confirmed from the qRT-PCR results (Fig. 2). After DESeq analysis, no differences in the relative expression levels of any exon 4 or 6 variants between unwounded and *E. coli*-challenged treatments at 6 hours (exon 4: all q-values $\geq$ 0.44; exon 6: all q-values $\geq$ 0.52) or 30 hours (exon 4: all q-values $\geq$ 0.76; exon 6: all q-values $\geq$ 0.96) were detected. Thus, no differential exon 4- or 6-expression was detected in this study using DESeq except exons 4.2 and 4.9, which were significantly lower in Ringers-wounded *Wolbachia*-infected flies at 30 hours, and exon 4.9, which was significantly lower in flies exposed to *M. luteus* at 6 hours compared to sterile-wounded flies. Following MCMCglmm analysis, an effect of *M. luteus* at 6 hours was seen on exon 4.9 in *Wolbachia*-cleared flies was seen, however, no other significant effects were revealed.

Finally, I tested if patterns of expression of exon 4-6 combinations were apparent between treatments. Figure 6 shows a heat map of the relative expression of exon 4-6 combinations of which all represented an abundance of over 0.5% in at least one treatment in the study.

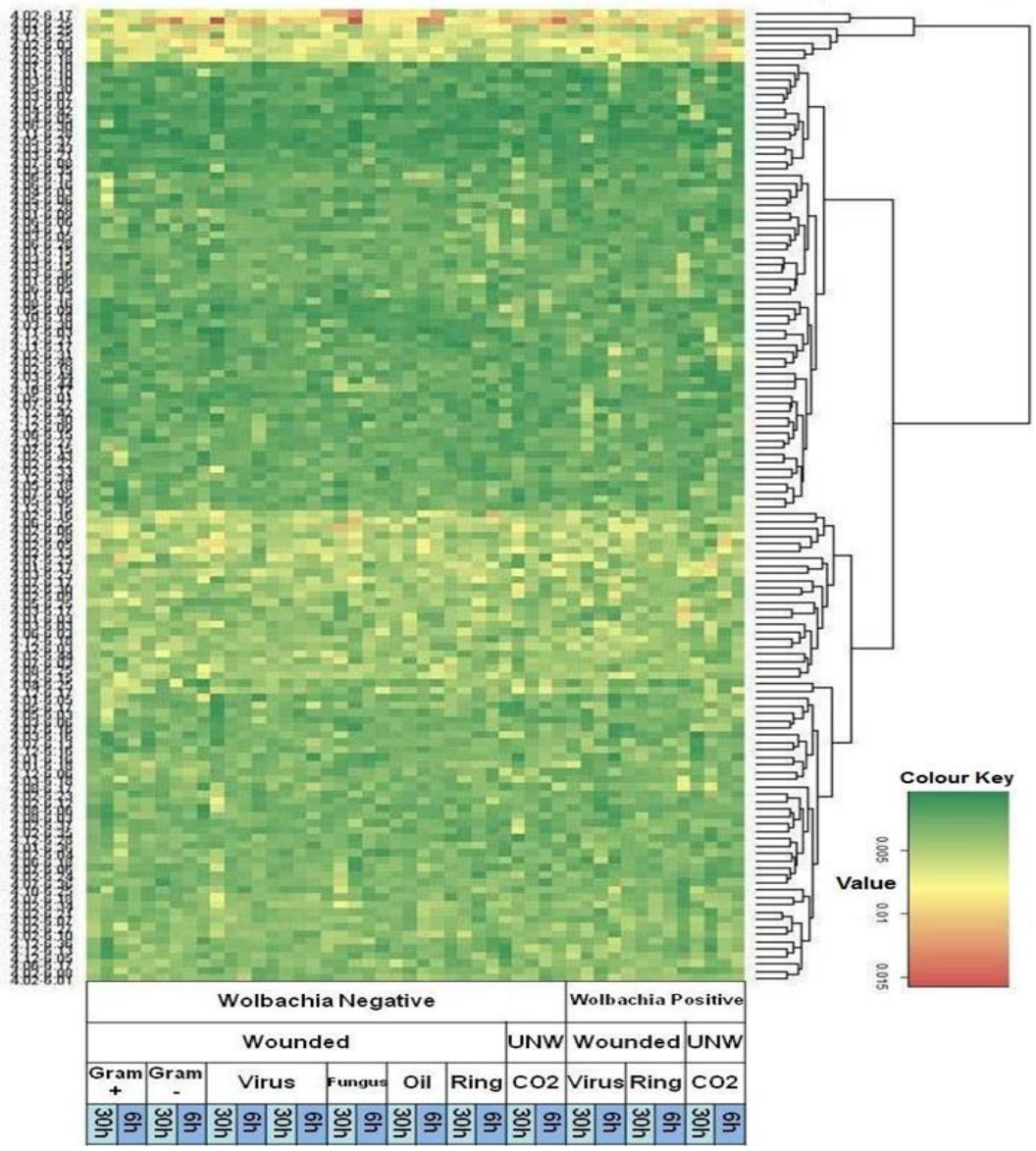


Figure 6: **Heat map representation of relative expression of exon 4-6 combinations of which each represented a minimum abundance of 0.5% in at least one treatment in this study.** Patterns of expression for exon 4-6 combinations across treatments were evident. For example, transcripts containing the combination 4.2-6.25 were the most abundantly expressed in most samples across all treatments. No patterns were obvious between treatments. Each time point included 2 biological replicates, and each cell represents the proportional expression of a given combination in a single sample. Dendrograms result from hierarchical clustering of exon combination-expression pattern using Euclidean distance measures. **Key**: *Wolbachia* Positive: flies harbouring *Wolbachia, Wolbachia* Negative: flies cleared of *Wolbachia,* UNW: Unwounded, CO2: carbon dioxide anaesthetised, Ring: Ringers solution, Oil: oil control, Fungus: *B. bassiana*, Virus: DCV and PiGV, Gram -: *E. coli*, Gram +: *M. luteus*, 6h: 6 hour time point, 30h: 30 hour time point.

As with individual exons, visual inspection suggested that patterns were evident across treatments. For example, transcripts containing the combination 4.2-6.25 were the most abundantly expressed in most samples across treatments. However, no patterns of expression were apparent between treatments. After applying the same methods of analysis as for individual exons (see above), no significant differences of expression of exon 4-6 combinations between any treatments in this study were detected. Thus, visual inspection of relative expression and analysis with DESeq and MCMCglmm revealed no effects of challenge between treatments.

3.3.6. Comparing the data with an independent cloning and sequencing assay

The study also included duplicated treatments where Illumina sequencing was carried out on a small number of samples from separate PCR reactions of the same cDNA samples. These duplicated treatments also showed dissimilar patterns of exon variant expression (not shown). Therefore, to investigate the reliability of the assay, I also carried out a separate cloning and sequencing assay to determine whether the results were comparable. I selected two treatments which showed the most contrasting differences in expression. As such, I cloned and sequenced 100 cDNAs for a DCV treatment and 100 cDNAs for a PiGV treatment. Figure 7 shows bar charts of the proportions of exon 4 and 6 variants for both treatments and for both sequencing assays.

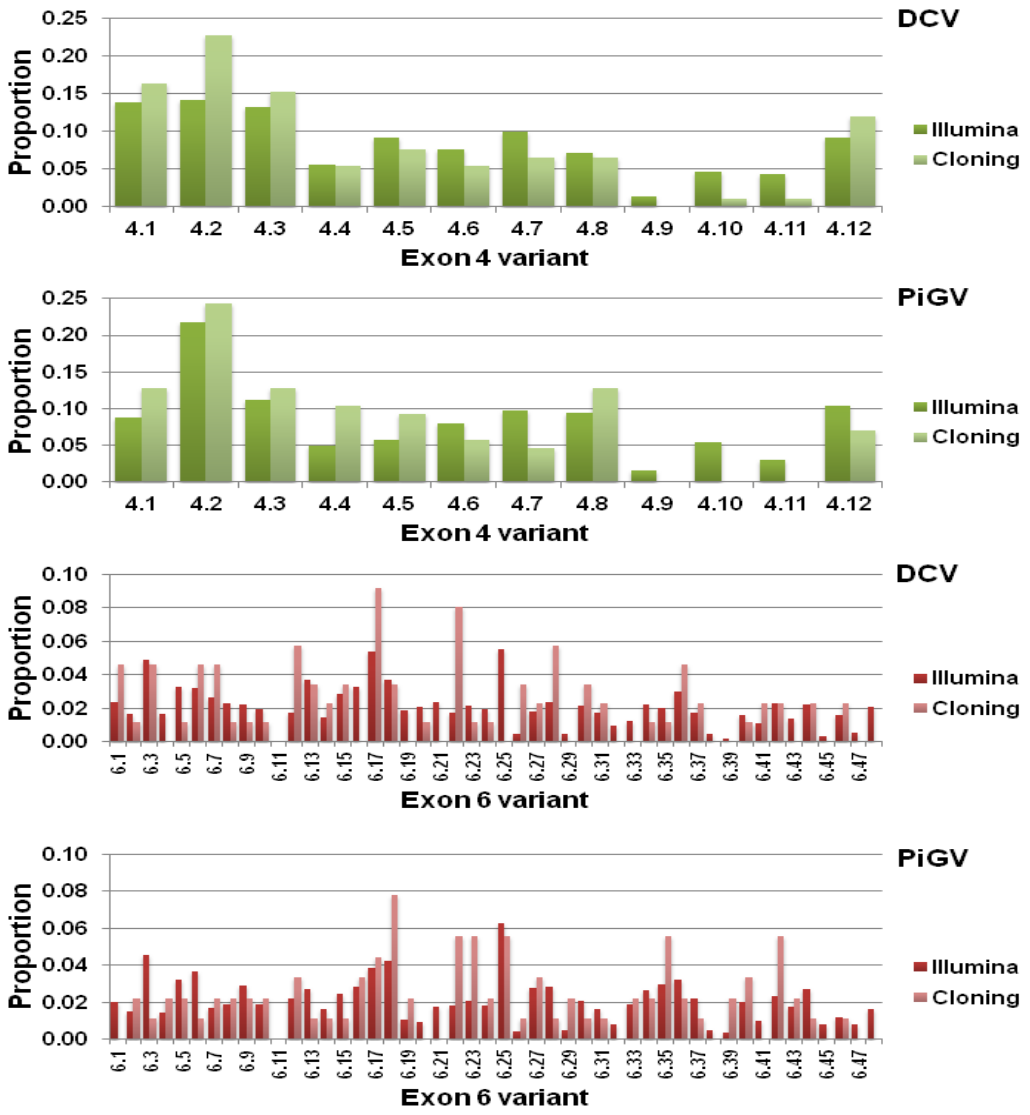


Figure 7: Bar chart representation comparing proportions of exon 4 and 6 variant expression between Illumina sequencing results and cloning and sequencing assay results in a DCV treatment and a PiGV treatment. Mixed results showed no difference between methods for exon 4-expression in the DCV treatment (Fisher's Exact Test, $p = 0.298$), whereas exon 6 in the DCV treatment showed a significant difference between assays ($p = 0.007$) possibly driven largely by the higher detected expression of exon 6.22 in the cloning and sequencing assay. Opposite results were seen for PiGV. Exon 6-expression showed no difference between assays ($p = 0.145$), however exon 4-expression was significantly different between the methods ($p = 0.012$), possibly driven largely by a higher detected expression of exon 4.4 in the cloning and sequencing assay.

The results showed no difference in exon 4-expression in the DCV treatment between the two methods (Fisher's Exact Test, p-value = 0.298), however exon 6-expression in the DCV treatment showed a statistically significant difference between the methods (p = 0.007). Opposite results were seen for PiGV. The proportion of exon 4 variants expressed between the two assays appeared to be different (p = 0.012), while no differences were seen in exon 6-expression in the PiGV treatment (p = 0.145).

3.3.7. qRT-PCR confirmed harvest time-points were appropriate

To investigate whether I chose appropriate time-points to harvest flies, I carried out a separate qRT-PCR experiment infecting flies with different types of bacteria and found up-regulation of classical immune markers Drosomycin and Diptericin at 6 hours post-exposure to *M. luteus* and *E. coli*, respectively (see appendix, figure S8). More specifically, levels of Diptericin had peaked between 4-8 hours after exposure to *E. coli* and had dropped to pre-exposure levels between 24-48 hours (Fig. S8). Drosomycin levels in response to *M. luteus* increased steadily from the time of exposure throughout the time-course and levels remained up-regulated after 30 hours, peaking around 48 hours (Fig. S8). These data indicated that the time points chosen to harvest flies were reasonable to test Dscam expression in response to immune challenge.

3.4. Discussion

3.4.1. Discerning and quantifying constitutive Dscam expression in Drosophila

Here, the Illumina-based sequencing assay provided millions of reads of which were assigned to many *Dscam* exon variants across different treatments in this study. The method provided a substantial amount of data on Dscam expression in *Drosophila* in general. In

particular, it discerned and quantified non-random expression of both exons 4 and 6 in unchallenged *D. melanogaster* (Fig. 3). For example, chi-square analysis confirmed that expression of individual exon variants strongly deviated from even expression, and MCMCglmm analysis showed that much of the variance in each of the implemented models was explained by expression of exons 4 and 6 (see Tables 1 and 2). Thus, the data demonstrated that exon variant choice in *Drosophila* is non-random.

The method was also able to detect the combinatorial splicing of exons 4 and 6 in *Drosophila*. The data suggests that most exon 4-6 combinations are expressed randomly (Fig. 4). However, some combinations seemed to be expressed in a non-random fashion. In other words, individual variants were combining at frequencies other than those expected as estimated from their individual frequencies. Examination of individual replicates showed similarities between exon-exon combination-expression suggesting that any effects introduced by technical variation or PCR-based stochasticity were minimal. These phenomena are discussed further below.

Two allelic variations of exons 6.12 and 6.39 were discovered, both different at one base each, and the DNA sequence variations could be the result of SNPs. No allelic variation was detected for exon 4, possibly indicative of different selective pressures acting upon each exon cluster (see Chapter 1: General Introduction). Thus, the deep sequencing assay used in this study was effective in detecting allelic variation between alternatively spliced exon variants of *Dscam* and appears to have discovered non-classical alleles which would not have been detected using some other commonly used methods (see Introduction). The method is also capable of detecting novel exons.

Exon 4- and 6-skipping was also detected in this study. It is thought that missed exons generate truncated and non-functional proteins (Schmucker and Chen 2009; Tazi, Bakkour et

al. 2009), and as such, may represent failure of the splicing machinery or could be part of some regulatory mechanism. Here, the frequency of exon 6-skipping appeared to occur more than exon 4-skipping by over two orders of magnitude. However, as the forward primers used here were seated in conserved exon 3 and overlapped variable exon 4 by one base, a true estimate of exon 4-skipping-frequency was not possible. Thus, I found that both exon 4- and exon 6-skipping occurs in *Drosophila*, but the frequency of exon-skipping could not be directly compared between the exons.

Lastly, exon 6.11 was entirely absent, an observation seen in earlier studies (Schmucker, Clemens et al. 2000; Hummel, Vasconcelos et al. 2003; Neves, Zucker et al. 2004; Watson, Puttmann-Holgado et al. 2005). The reason for the absence of expression of exon 6.11 remains unknown but lack of necessary complementarity between the docking site and selector sequence may result in expression failure (see Graveley 2005). Future genomic analysis of *Dscam* may help explain why exon 6.11 is the only exon 6 variant that is not expressed in *D. melanogaster*.

3.4.2. Dscam expression in response to wounding or pathogens

As pathogen-specific effects on Dscam expression may have been too subtle for cloning and sequencing to detect (see Chapter 2), a more powerful method was required. However, the Illumina-based sequencing assay did not detect any definitive effects of treatment on the expression of exon 4, exon 6, or exon 4-6 combinations in whole adult *Drosophila*. Initial visual inspection suggested that pathogen-induced effects, should they exist, must be small or difficult to detect (see figures 5 and 6).

I found only weak evidence of wounding effects on the expression of *Dscam* in *Drosophila*. Specifically, only DESeq analysis detected differential expression of some exon 4 variants in treatments in response to wounding. In *Wolbachia*-positive flies, exons 4.2 and 4.9 appear to

be significantly lower in flies receiving sterile-wounding harvested at 30 hours. This was not seen in *Wolbachia*-cleared flies. However, given that *Wolbachia*-infection is highly prevalent in most populations of *D. melanogaster* (Verspoor and Haddrill 2011), and that absence of *Wolbachia* appears to be the less natural state and may affect the efficiency of the immune response (Fig. 2), differences between the treatments are not necessarily unexpected. Thus, the very small but detectable effect of wounding on Dscam expression in *Drosophila* could be real. However, the results of different and complementary analyses were not consistent.

Pathogen-exposure has been shown to affect Dscam splicing in *Anopheles* (Dong, Taylor et al. 2006) and *Pacifastacus* (Watthanasurorot, Jiravanichpaisal et al. 2011). Here, indices of expression diversity were no different between sterile-wounded or bacteria-challenged treatments. However, both DESeq and MCMCglmm analysis detected differential expression of exon 4.9 between sterile-wounded and *M. luteus*-exposed flies at 6 hours, suggesting this exon may play a part in the immune response to the bacterium. Thus, various statistical analyses of Dscam expression in *Drosophila* found only weak evidence of pathogen-specific responding of exon 4, and no evidence of pathogen-specific responding of exon 6.

The pathogen-specific expression of *Dscam* reported in previous studies was only weakly supported here in *Drosophila*. Albeit comparing very different host organisms, this phenomenon could have been expected in *Drosophila Dscam* based on its implicated role in phagocytosis and the hyper-diversity of isoforms discovered in the fly's immune cells (Watson, Puttmann-Holgado et al. 2005). However, no strong effects were detected. It is possible the assay did not work very well, resulting in a failure to detect strong effects of treatment on Dscam expression. For instance, stoichiometry of the PCR could have resulted in random errors in amplification in early cycles being propagated in subsequent cycles. This can happen due to the significant skewed ratio between small amounts of template DNA and

excess amounts of PCR reagents. The template DNA in this case also comprises of highly diverse, and subsequently very rare expressed Dscam exon transcripts, and PCR stochasticity in early rounds may make it very difficult to obtain accurate estimates of their relative abundances. Also spurious targets which are not the preferred targets can be amplified in the early stages of PCR, and the amplification of these extraneous products continues exponentially (Ruano, Brash et al. 1991). Furthermore, primer targets with identical annealing sites but with slightly different fragment lengths and base compositions could have been impaired by thermodynamic or kinetic effects which may render one variant a preferred template over another (Ruano, Brash et al. 1991). Effects which reside at the level of RNA extraction and/or cDNA synthesis cannot be excluded either.

Appropriately, the mixed results from the cloning and sequencing assay (Fig. 7) encouraged the comparison between the Illumina sequencing results of this study with the expression data obtained from a past microarray analysis of *Dscam* (see Neves, Zucker et al. 2004). The microarray study was carried out on the same line of *D. melanogaster* as the present study, and fly-aging and conditions were comparable between the experiments. Mean proportions of expression of exon 4 and 6 variants of both studies was subjected to correlation and dependence testing. As such, Pearson's correlation coefficients were calculated for exons 4 and 6. Figure 8 compares the relative expression of exon 4 and exon 6 variants between the Illumina-based sequencing assay of the present study and a microarray analysis of *Dscam* in *D. melanogaster* (Neves, Zucker et al. 2004).

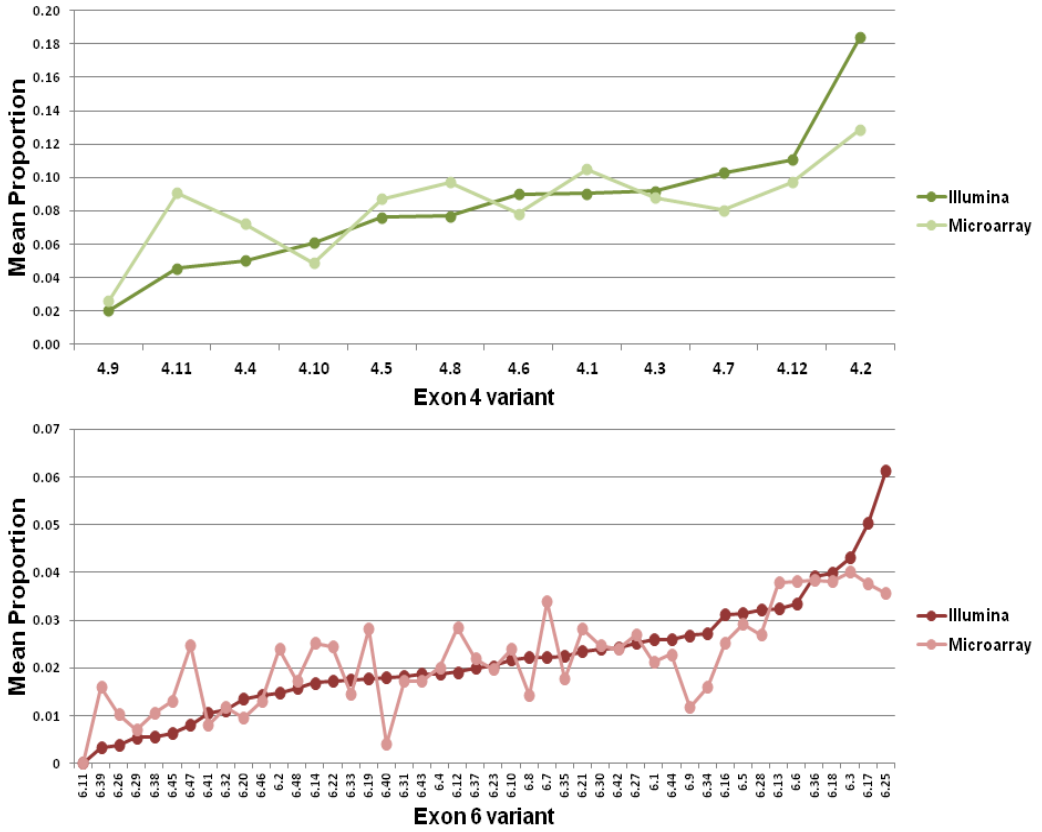


Figure 8: Scatter plot representations of relative expression levels of exons 4 (top) and 6 (bottom) between the Illumina-based sequencing results of this study and data from an independent microarray study (Neves, Zucker et al. 2004). Pearson's correlation coefficients were calculated for each exon using their mean relative proportions (exon 4, r = 0.79, exon 6, r = 0.77), and suggested a positive association between the data of the two independent studies.

I found a positive association between the results of these studies (Pearson's correlation coefficient: exon 4, r = 0.79, exon 6, r = 0.77). Thus, similarities between the Illumina sequencing data and those of an independent microarray study were encouraging. While some differences were present, these are to be expected as they were totally independent experiments using different insects and carried out by different people using two completely dissimilar methods. Indeed, differences could be attributed to the fact that sequencing offers much higher sensitivity and dynamic range, and compared to microarray assays, suffers from lower technical variation (Oshlack, Robinson et al. 2010). Interestingly, I noticed that the

only two exons which were discovered to have allelic variation in the data set of the present study (exons 6.12 and 6.39, see above), also showed notable higher relative expression levels in the independent microarray study (Fig. 8). Results could differ between the methods due to the presence of SNPs, or if the microarrays did not match the reference genome. With regard to the cloning and sequencing results (Fig. 7), it was possible that not enough observations were taken, subsequently making it difficult to directly compare with high-throughput sequencing data. Nevertheless, while some data did not correlate well between the sequencing assays, some data correlated strongly. Thus, the comparison of the Illumina-based sequencing data with those of the independent cloning and sequencing assay and the independent microarray study lead me to propose that the method was not significantly compromised by stoichiometric effects of early-round PCR, and that the high-throughput sequencing assay used here was reliable and could be adopted for future study of *Dscam* gene expression. Nevertheless, the pooling of several independent PCRs for each treatment in future comparable studies may help reduce the likelihood of spurious amplification.

It is also possible that the diversity of such an exceptionally diverse splicing gene is too high to detect effects of pathogen-exposure on Dscam expression in *Drosophila* even at this depth of coverage. This may also have been a consequence of having many (~25) whole flies in each treatment. High levels of *Dscam* present in the insect's nervous system may contribute too much noise for a much smaller pathogen-effect to be detected should it exist. The solution to this may be the separation and comparison of tissues (e.g. heads, haemolymph and rest of body) to minimise the contribution of nervous-system Dscam. Investigation of tissue-specific Dscam expression may help detect and focus upon any pathogen-effects on *Dscam* splicing should they occur, as well as offer more insight into gene variant abundance and expression in different tissues in general.

Furthermore, pathogen-specific splicing of *Dscam* may simply not happen in *Drosophila*, or indeed pathogen-specific splicing in response to the challenges used in this study may not occur in *Drosophila*. In this study, the use of *natural* pathogens had no detectable effects of challenge on the expression of *Dscam*. However, the data from the qRT-PCR assays for both DCV and *B. bassiana* were unconvincing (Fig. 2). Therefore, that no effect on Dscam expression was detected for these pathogen-exposures may be due to the failure in induce a sufficient corresponding immune response in the flies. *E. coli* in particular produced more convincing effects of exposure (namely, highly increased levels of Diptericin, see figure 2) and although it is not a natural pathogen of *Drosophila*, it has been previously shown that recombinant Dscam isoforms may bind specifically to *E. coli* in the fly (Watson, Puttmann-Holgado et al. 2005). Furthermore, *E. coli* has been reported to induce pathogen-specific effects on Dscam expression in other species (Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). Therefore, a pathogen-specific *Drosophila Dscam* response to *E. coli* could have been expected.

Finally, in my previous work on *Anopheles Dscam* in Chapter 2, I found a statistically significant difference in expression diversity between unexposed and exposed treatments. Although it was the combined diversity of exons 4 and 6 that drove the pattern, the relationship between exon 4 expression-diversity and parasite diversity was close to significant in two independent experiments (field, $p = 0.081$; lab, $p = 0.087$). Thus, observations from two completely independent studies in *Anopheles*, and another here in *Drosophila* using three different methods of analysis, suggest that an exon 4-by-exposure effect in *Anopheles* and an exon 4-by-exposure effect in *Drosophila* could be real. It is interesting that any effects of Dscam expression detected, however small, are always associated with exon 4, and not exon 6.

3.5. Conclusion

In the present study, I found no supporting evidence of a strong effect of treatment on the expression of exon 4, exon 6, or exon 4-6 combinations in whole adult *Drosophila*. However, results from this study and those of Chapter 2 suggest that immune-challenge may have a small effect on Dscam exon 4-splicing. As such, a natural next step in the study of Dscam expression would be to investigate the tissue-specific expression of exon variants in response to strong and reliable immune elicitors.

Chapter 4: Tissue-specific Dscam gene expression in response to gram-negative and gram-positive immune elicitors

This chapter was written with comments from Darren Obbard and Tom Little. Statistical analyses were done in collaboration with Darren Obbard. The quality trimming, counting and sorting of raw Illumina sequencing reads was done using a custom R pipeline written by Darren Obbard. Illumina library construction, sequencing and base-calling were performed by the GenePool Sequencing Facility (University of Edinburgh).

4.1. Introduction

The Dscam locus contains many different versions of exons which code for variable regions of three extracellular immunoglobulin domains of the Dscam axon guidance receptor (Schmucker, Clemens et al. 2000) (see Chapter 1, General Introduction). Mutually exclusive alternative splicing of these exon variants provides high functional diversity which is essential for the development of the nervous system in *Drosophila* (Schmucker, Clemens et al. 2000; Chen, Kondo et al. 2006; Hattori, Millard et al. 2008). Dscam has also been implicated in pathogen-response in several species and may play an important part in the phagocytosis of pathogens after infection (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). For instance, RNAi-knockdown of *Dscam* in both *Drosophila* and *Anopheles* results in a significantly reduced phagocytic capacity (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006). Phagocytosis in insects is carried out by haemocytes, circulating cells in the haemolymph (Meister 2004), and a high diversity of Dscam isoform-expression has been found in *Drosophila* haemocytes in addition to the nervous system, indicating an importance of Dscam expression in different tissues (Watson, Puttmann-Holgado et al. 2005).

Dscam splicing appears to be spatially regulated. For example, by comparing exon 4-expression in different tissues (antennae, heads, wings and legs) of *Drosophila*, Celotto and Graveley (2001) found that the relative frequency of exon 4.2-usage was significantly higher in legs and wings than in heads or antennae. Independently, using custom-made oligo-arrays with probes for all alternatively spliced exons, Watson et al (2005) demonstrated that variants of both exons 6 and 9 showed restricted usage in haemocytes and the fat body compared to brain tissue, while exon 4 variants were expressed in all tissue types with moderate to high expression. Furthermore, microarray analysis of Dscam expression between brain and developing eye disks found only very small differences in the relative

expression of exons 4 and 6, but a broad yet distinctive spectrum of isoforms were expressed between the tissue types due to diverse exon 9-inclusion (Neves, Zucker et al. 2004). *Dscam* may be regulated in a tissue-specific manner to enable neurons in different tissues to express different suites of Dscam receptors in order to direct their axons and dendrites to specific 'addresses' (Celotto and Graveley 2001).

Tissue-specific expression of Dscam variants has also been discovered in other species. In the crustacean, *Daphnia*, fewer variants of all variable exons encoding extracellular domains are expressed in haemocytes than in brain tissue (Brites, McTaggart et al. 2008). Moreover, some exon-exon combinations in *Daphnia* are 'preferred' in the brain compared with the haemocytes (Brites, McTaggart et al. 2008). Finally, tissue-specific expression of Dscam mRNA has been reported in the shrimp, *Litopenaeus vannamei*. Here, Dscam transcripts were highly expressed in heart and lymphoid tissue, in lower abundance in intestinal tissue and haemocytes, and not detected at all in the stomach or hepatopancreas (Chou, Chang et al. 2009).

In Chapter 3 I used Illumina sequencing technology to investigate Dscam gene expression in whole adult *Drosophila* exposed to multiple pathogen-types. However, it is possible that Dscam expression in the nervous system of *Drosophila* may have obscured smaller pathogen-effects on the expression of Dscam exon variants. Here, I recorded the expression of exons 4 and 6, and combinations thereof, in different tissues (head, hemolymph, and the remainder of the body including legs and wings, hereafter 'body') in adult *Drosophila* exposed to different immune elicitors. Separation of heads was chosen to isolate as much as possible the contribution of nervous system expression of *Dscam* from the rest of the insect, but also to compare expression in heads of flies exposed to different immune challenges. Dscam expression in haemolymph was chosen as it contains the cells which confer cell-

mediated immunity in invertebrates (Stuart and Ezekowitz 2008). The bacteria *Escherichia coli* and *Micrococcus luteus* were chosen as immune elicitors.

4.2. Methods

4.2.1. Fly rearing and immune challenge

Fly rearing was identical to Chapter 3. *Escherichia coli* (DH10B strain, Invitrogen) and *M. luteus* (LZB055 strain, Blades Biological Ltd, Edenbridge, UK) were used as immune elicitors. Bacterial preparation was the same as Chapter 3. Immune challenges were: unwounded control (anaesthetised only), wounded control (isotonic Ringers solution for *Drosophila*; 182mM KCl, 46mM NaCl, 3mM $CaCl_2$, 10mM Tris-Cl), *E. coli* (gram-negative bacterium), and *M. luteus* (gram-positive bacterium).

Flies were anaesthetised with CO_2, and then pierced in the side of the thorax with a pin (diameter ~ 0.05mm) that had been dipped in either Ringers solution or bacterial suspension. Flies were then placed in vials containing a sugar-agar medium and kept at 18°C. After 6 hours, flies were anaesthetised with CO_2, and heads (with antennae) were removed and immediately placed into cold TRIzol reagent (Invitrogen). To extract haemolymph (containing haemocytes), the thorax was then pierced with a pin (diameter ~ 0.2mm), and flies were placed into chilled Eppendorf tubes punctured at the bottom using a hypodermic needle (diameter ~ 0.8mm). They were then immediately centrifuged at 9,000g for 5 minutes at 4°C to drain hemolymph directly into cold TRIzol reagent in a second Eppendorf tube supported below. Finally, haemolymph-depleted bodies (including legs and wings) were placed into cold TRIzol reagent in separate tubes. The separated tissues immersed in TRIzol reagent were immediately placed in liquid nitrogen and shortly after homogenised and removed to -80°C pending RNA extraction. A total of 25 individual flies were used for each

of 4 treatments (unwounded, sterile-wounded, gram-negative, gram-positive), and the experiment was replicated 6 times. A total of 600 flies were used in this study.

4.2.2. RNA isolation, reverse transcription, PCR and Illumina Sequencing

Methods were identical to Chapter 3. All approaches using PCR to amplify cDNA from a sample can be affected by PCR stochasticity. The early rounds of PCR constitute a 'screening phase' in which particular cDNA transcripts are chosen for amplification (Ruano, Brash et al. 1991). Thus, targets are selected for amplification prior to the amplification beginning. When attempting to determine accurate relative quantities of Dscam exon variants within a sample, stoichiometry of the PCR may distort the true pattern. As such, to reduce the possible effects of PCR stochasticity in this study, PCR of each cDNA sample was repeated 5 times and pooled together.

4.2.3. Data analysis

Raw data handling

Raw data was handled as in Chapter 3.

Statistical Analyses

Analysis of constitutive exon expression within different tissues: I tested for a departure from even expression in different tissue types in *Drosophila* using different methods of analysis. Pearson's chi-square (χ^2) tests were carried out on the mean raw counts of exons 4 and 6 of unchallenged *Drosophila* to determine whether the frequency distribution of each exon variant was consistent with a random expression model (i.e. all exon variants are equally likely to be expressed; expectation calculated as 1/n variants at a given exon cluster). Exon 4-6 combinations in unchallenged *Drosophila* were tested for a departure from even expression using Fisher's Exact Tests on raw count data for each treatment due to low counts for some combinations. Generalised linear mixed models using MCMCglmm (Hadfield

2009) were also implemented to reveal non-random constitutive expression in different tissues by providing estimates of the proportion of variance explained by exons 4 and 6. See Chapter 3 Methods for description of MCMCglmm package. All analyses were handled in R (R Development Core Team 2011).

Comparisons of constitutive exon expression between different tissues: I compared the constitutive expression of exons 4 and 6 between tissues to test for tissue-specific Dscam expression using three different methods of complementary analysis. DESeq (Anders and Huber 2010) was used to test for tissue-specific expression of exons 4 and 6 by testing the number of reads for each Dscam exon variant for being differential between tissues. Here, gene variants were treated as equivalent to genes. DESeq corrects for multiple testing using Benjamini-Hochberg's FDR (Benjamini and Hochberg 1995), and here FDR was controlled at 5% (q = 0.05). See Chapter 3 Methods for description of DESeq package. MCMCglmm was used to infer tissue-specific expression by providing estimates of the proportion of variance explained by exon-by-tissue effects in the model. Additionally, diversity of expression between tissues of unchallenged *Drosophila* was compared using Simpson's Index of diversity (1-D) (Simpson 1949) and analysed using traditional one-way ANOVA of four treatments, each with six individual observations (biological replicates). All analyses were handled in R (R Development Core Team 2011).

Analysis of infection-responsive Dscam splicing in Drosophila: I used the same three approaches to detect infection-responsive splicing as I did to detect tissue-specific splicing (see above). As *Dscam* can produce many thousands of isoforms through alternative splicing an estimate of diversity can be useful measure of the overall Dscam transcript-range present under different conditions. As such, expression diversity was estimated using Simpson's Index of diversity (1-D). A general linear model (GLM) was fitted to indices of diversity to test for effects of tissue, pathogen, and an interaction between both as: (1-D) ~ tissue +

pathogen + tissue*pathogen. I used two complementary methods to examine differential expression of exons 4 and 6 in response to immune challenge. Firstly, I used MCMCglmm to infer the effect of wounding or pathogen-exposure on Dscam expression patterns. Specifically, I compared Dscam expression between (1) unwounded and Ringers-wounded treatments (i.e. to investigate any wounding effects on Dscam expression and in different tissues), and (2) different pathogen-exposed treatments compared to sterile-wounded treatments (i.e. to investigate any effects of pathogen-exposure and in different tissues). I fitted models to the count data to test for an interaction between Dscam expression and experimental conditions. To test for effects of wounding, the following model was fitted to exon read counts: reads ~ tissue + wounding + replicate + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: tissue + exon 4: wounding + exon 6: tissue + exon 6: wounding + exon 4: tissue: wounding + exon 6: tissue: wounding + exon 4-6: tissue + exon 4-6: wounding + exon 4-6: tissue: wounding. Tissue and wounding were treated as fixed effects and pathogen-exposed treatments were excluded from this analysis. To test for effects of pathogen exposure, I fitted the model: reads ~ tissue + pathogen + replicate + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: tissue + exon 4: pathogen + exon 6: tissue + exon 6: pathogen + exon 4: tissue: pathogen + exon 6: tissue: pathogen + exon 4-6: tissue + exon 4-6: pathogen + exon 4-6: tissue: pathogen. Tissue and pathogen were treated as fixed effects and unwounded treatments were excluded from this analysis. Relative expression was transformed as cube-root of percentages (expressed as reads per million) to ensure normally distributed residuals (see appendix figure S9 for Q-Q plot of distribution before and after transformation). Two runs of 1,000,000 MCMC iterations were completed for each model. Each run was sampled every 100 steps and the first 10% of iterations was discarded as burn-in. Secondly, DESeq was also used to detect infection-induced differentially expressed variants of exons 4 and 6, and combinations thereof. Here, I tested for differential expression of exon variants between (1) unchallenged and sterile-wounded treatments to test for an effect of wounding, (2) sterile-wounded and septic-wounded treatments to test for an effect

of pathogen-exposure, and (3) comparisons 1-2 above but testing the expression of exon 4-6 combinations. FDR was controlled at 5% (q = 0.05). All analyses were handled in R (R Development Core Team 2011).

Correlation and dependence: I also tested whether proportional expression of exon 4 and 6 variants correlated between the data of the present study and those of whole flies (Chapter 3) using Spearman's rank correlation (r_s). Correlation analyses were handled in R (R Development Core Team 2011).

4.3. Results

4.3.1. Illumina sequencing assay output

Data received from the sequencing facility had 10,214,702 read pairs. After quality trimming (see Methods), I had 4,722,048 reads which were still available in pairs of reads at least 50nt long. Sequences without a valid MID-tag or affected by mis-priming or template switching were then discarded and a total of 1,469,404 million reads remained and were attributable to exons 4 and 6, and combinations thereof. The mean number of high quality reads from each PCR primer-pair was 20,408 (sequencing-depth ranged from 1,593 up to 60,218).

4.3.2. Detecting differences in constitutive exon variant expression in different tissues

I used Pearson's chi-square and MCMCglmm analyses to test for uneven expression of exons 4 and 6 in different tissues of unchallenged flies. I found expression patterns in all tissues to be very similar to those found in whole flies (see Chapter 3, Fig. 3). Significant differences in expression between variants of both exons 4 and 6 were detected in all three tissue types (exon 4 heads: $\chi^2 = 6394.42$, DF = 11, p = <0.0001; exon 4 bodies: $\chi^2 = 1875.59$,

DF = 11, p = <0.0001; exon 4 haemolymph: χ^2 = 1728.10, DF = 11, p = <0.0001; exon 6 heads: χ^2 = 30015.2, DF = 47, p = <0.0001; exon 6 bodies: χ^2 = 5327.09, DF = 47, p = <0.0001; exon 6 haemolymph: χ^2 = 7586.46, DF = 47, p = <0.0001) (Fig. 1). Using MCMCglmm, I found significant effects of exon 4 and exon 6-expression, indicating non-random expression of both exons. Specifically, exon 4 contributed around 53% of the total variance explained by random effects in the model, while exon 6 explained around 38% (Table 1). Thus, both methods of analysis showed that variants of exons 4 and 6 were expressed non-randomly, as seen previously in whole flies (Chapter 3).

4.3.3. Constitutive exon variant expression between different tissues

I used DESeq, MCMCglmm and Simpson's Index of Diversity analyses to compare constitutive exon expression between different tissues in unchallenged *Drosophila*. Relative expression of some exon 4 variants appeared to be different between tissues. In other words, some exon 4 variants were more abundant in specific tissue types than other variants (Fig. 1). For example, exon 4.2 was significantly more abundant in bodies than in heads (DESeq, q = <0.001) or haemolymph (q = <0.001). In contrast, exon 4.2-expression was comparable between heads and haemolymph (q = 0.78). Exon 4.9-expression was significantly higher in heads than in bodies (q = <0.001) and in haemolymph (q = <0.001), but no difference in the relative abundance of exon 4.9 between bodies and haemolymph was detected (q = 0.38). Exon 4.11 was expressed at significantly higher levels in heads than in bodies (q = <0.001) or haemolymph (q = 0.007), and expressed significantly more in haemolymph than in bodies (q = <0.001). MCMCglmm analysis revealed significant exon 4-by-tissue effects and was consistent with DESeq analysis (exon 4.2-by-body effect, MCMCglmm, p = 0.002; exon

4.9-by-head effect, p = 0.007; exon 4.11-by-head effect, p = 0.05).

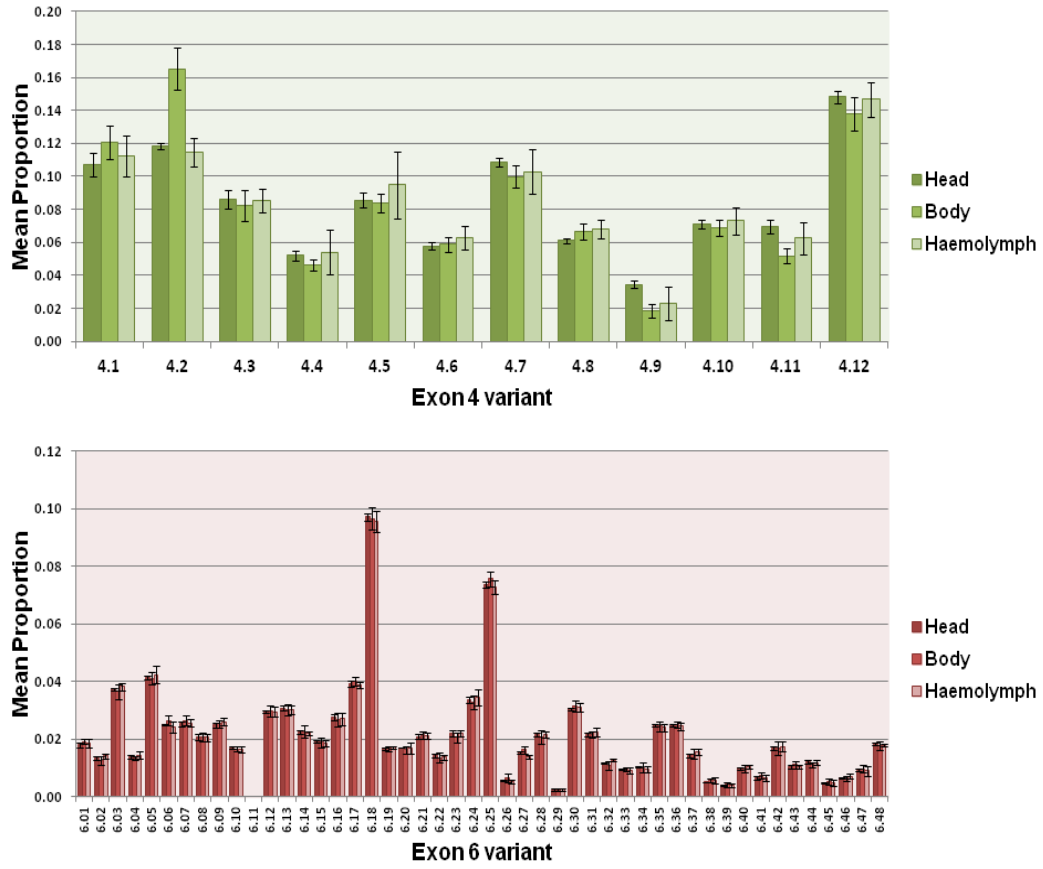


Figure 1: **Bar chart representing relative expression of each exon 4 and 6 variant in different tissues in unchallenged *D. melanogaster***. Exon 4 and 6 variants in heads, bodies and haemolymph were expressed in a non-random fashion. Furthermore, analysis showed that expression of individual exon 4 variants appeared to be tissue-specific. For example, exon 4.2 expression was significantly higher in bodies than in heads (DESeq, q = <0.001) or haemolymph (q = <0.001). Exon 4.11 was significantly higher in heads than in bodies (q = <0.001) or haemolymph (q = 0.007), while exon 4.9 was significantly higher in heads than in bodies (q = <0.001) or in haemolymph (q = <0.001). Moreover, significant exon 4-by-tissue effects were found following analysis with MCMCglmm (see text). In contrast, relative exon 6 variant expression did not differ between tissues. Error bars represent standard deviation calculated from six independent experiments.

Moreover, exon 4 expression-diversity was lower in bodies than in heads or haemolymph of unchallenged flies (One-way ANOVA; $F_{2,15} = 32.21$, $p = <0.001$). Thus, exon 4 variants appear to be expressed in a tissue-specific manner.

For exon 6 variants, the relative expression across tissues was remarkably similar (Fig. 1). For example, the most abundant variant found was exon 6.18 and was expressed in the same relative abundance in all tissues (between heads and bodies; DESeq, q = 1, heads and haemolymph; q = 0.99, and bodies and haemolymph; q = 0.83). Exon 6.25 was the second most abundant variant expressed in all tissues and no differences in expression of this variant were seen between tissues (between heads and bodies; q = 1, heads and haemolymph; q = 0.99, and bodies and haemolymph; q = 0.84). Analysis of the expression of the remaining variants of exon 6 found no differences in relative expression of any variant between heads and bodies (all q-values = 1), between heads and haemolymph (all q-values $\geq$ 0.99), or between bodies and haemolymph (all q-values $\geq$ 0.63). MCMCglmm analysis revealed no significant exon 6-by-tissue effects (Table 1). Moreover, diversity of exon 6-expression was no different between tissues (One-way ANOVA: $F_{2,15} = 0.89$, $p = 0.801$). Thus, the expression of exon 6 variants did not show the apparent tissue-specificity of exon 4 variants, and two exon 6 variants in particular (6.18 and 6.25) showed strikingly higher relative expression than any others (Fig. 1).

Notably, exon-skipping was detected in all tissues. Exon 6-skipping was detected at a frequency of $2.8x10^{-3}$ in heads, a frequency of $3.1x10^{-3}$ in bodies, and at a frequency of $2.6x10^{-3}$ in haemolymph. In this experiment, although exon 4-skipping was detected in all tissues, I could not measure the rate with any accuracy as the forward primers which were nested in constitutive exon 3 also overlapped variable exon 4 by one base. Thus, both exon 4- and exon 6-skipping was detected in all tissues of unwounded flies, and exon 6-skipping appeared to occur at a similar rate in all tissues.

4.3.4. Detecting normal combinatorial splicing in different tissues in unwounded Drosophila and determining whether splicing is random or non-random

I next compared the constitutive combinatorial splicing between the tissue types of unchallenged flies and attempted to determine whether splicing appeared to be random or non-random in each tissue in the absence of wounding or parasites. Figure 2 shows heat maps of relative expression levels of all possible exon 4-6 combinations, and as in whole flies (see Chapter 3), visual inspection suggests that the splicing of these two exons is close to random in all three tissue types. However, deviation from even expression was detected in a small number of treatments and was seen in all three tissue types (see Table S5). Moreover, MCMCglmm analysis revealed a very small but significant exon-exon combination-by-tissue effect, which explained 0.1% of the total variance in the model (Table 1), indicating that a small number of exon 4-6 combinations were not evenly expressed. Thus, as seen in whole flies (Chapter 3), in general, combinatorial splicing appears to be close to random in different tissues in the fly, but some exon-exon combinations were detected at non-random frequencies in heads, bodies and in haemolymph.

Overall, all possible exon 4-6 combinations were detected in all three tissue types except those requiring exon 6.11, resulting in a total of 564 combinations expressed in heads, in bodies and in haemolymph of *D. melanogaster*. Thus, high Dscam transcript diversity was present in unwounded flies which demonstrated a non-restricted expression for both studied exons in all tissue types studied.

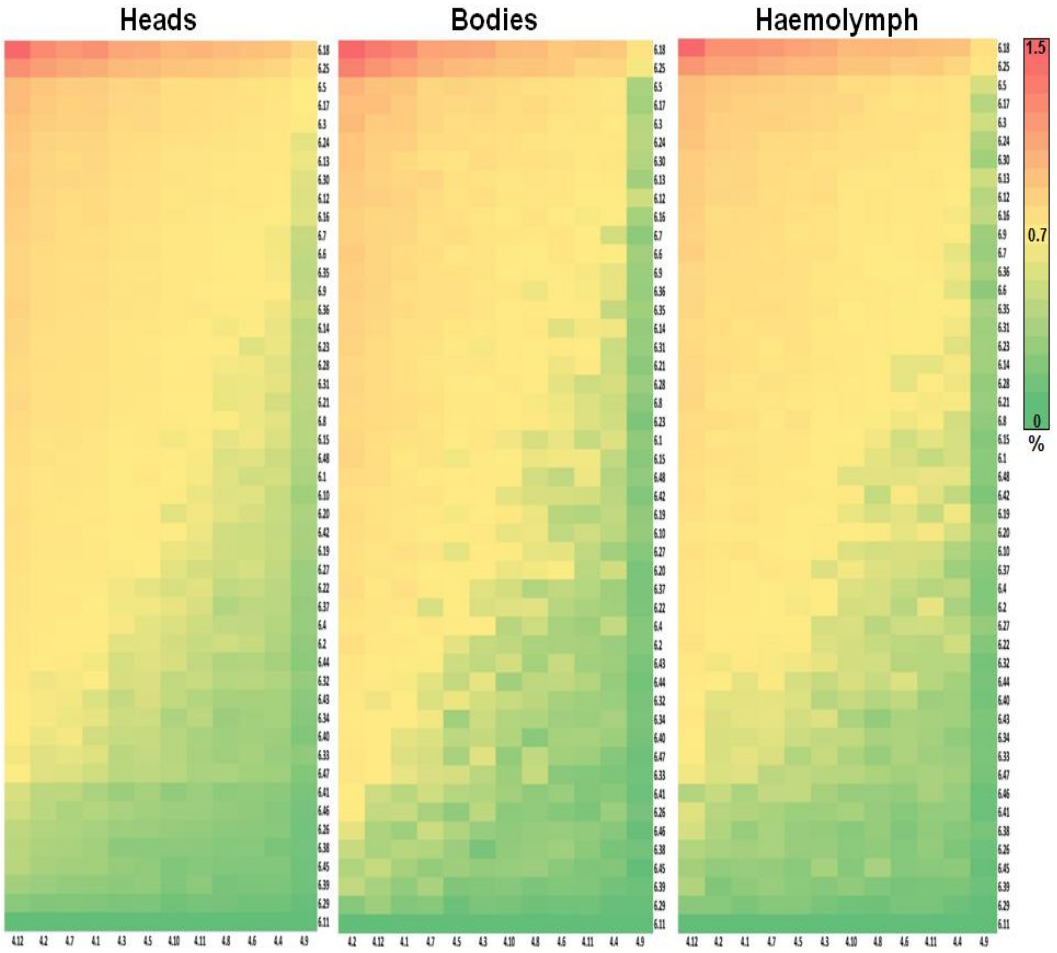


Figure 2: **Heat map representation of the relative expression of all possible exon 4-6 combinations in heads (left block), bodies (middle block), and haemolymph (right block) in unwounded *D. melanogaster***. Combinatorial splicing was close to random in all three tissues. For example, combination 4.10-6.28 found in heads (left block), and combination 4.2-6.9 found in haemolymph (right block), combined at frequencies expected based on the independent frequencies of each variant detected in these tissue types. However, some combinations appeared to show non-random association. For example, the most abundant combination in bodies was 4.2-6.18, which represented approximately 1.5% of all sequencing reads obtained from this tissue, was detected ~30% higher than would be expected in a random combinatorial splicing model based on independent frequencies of variants 4.2 and 6.18. Variants are in order of relative abundance for each tissue. Each cell represents the proportion of an exon 4-6 combination, calculated as the mean of biological replicates of unwounded treatments for each tissue (n=6).

4.3.5. Testing for differential Dscam expression in different tissues in response to different immune elicitors

I used Simpson's Index of Diversity, MCMCglmm and DESeq analyses to test whether Dscam expression was altered in response to immune challenges. Figure 3 shows bar charts of the mean Simpson's Index of Diversity values calculated for exon 4, exon 6, and exon 4-6 combinations for each tissue and each immune challenge (see Table S6 for diversity estimates for all treatments in the study).

There was a significant effect of tissue on the expression diversity of exon 4 ($F_{2,68}$ = 112.73, p = <0.001), where diversity was lowest in bodies. No effect of pathogen was found on exon 4 diversity ($F_{3,68}$ = 1.66, p = 0.184), and I found no significant interaction between tissue and pathogen ($F_{2,67}$ = 0.83, p = 0.549). For exon 6, no significant effects were found (tissue: $F_{2,68}$ = 0.21, p = 0.652; pathogen: $F_{2,68}$ = 0.98, p = 0.410; interaction: $F_{2,67}$ = 0.77, p = 0.598). Finally, a significant tissue-effect was found for exon 4-6 combination diversity ($F_{2,68}$ = 43.05, p = <0.001) where diversity was lowest in bodies, and no effect of pathogen ($F_{2,68}$ = 0.75, p = 0.526), or effect of interaction was seen ($F_{2,67}$ = 0.78, p = 0.589) (Fig. 3). Thus, while between-tissue differences were seen in expression diversity, no significant effects of pathogen-exposure on diversity were detected.

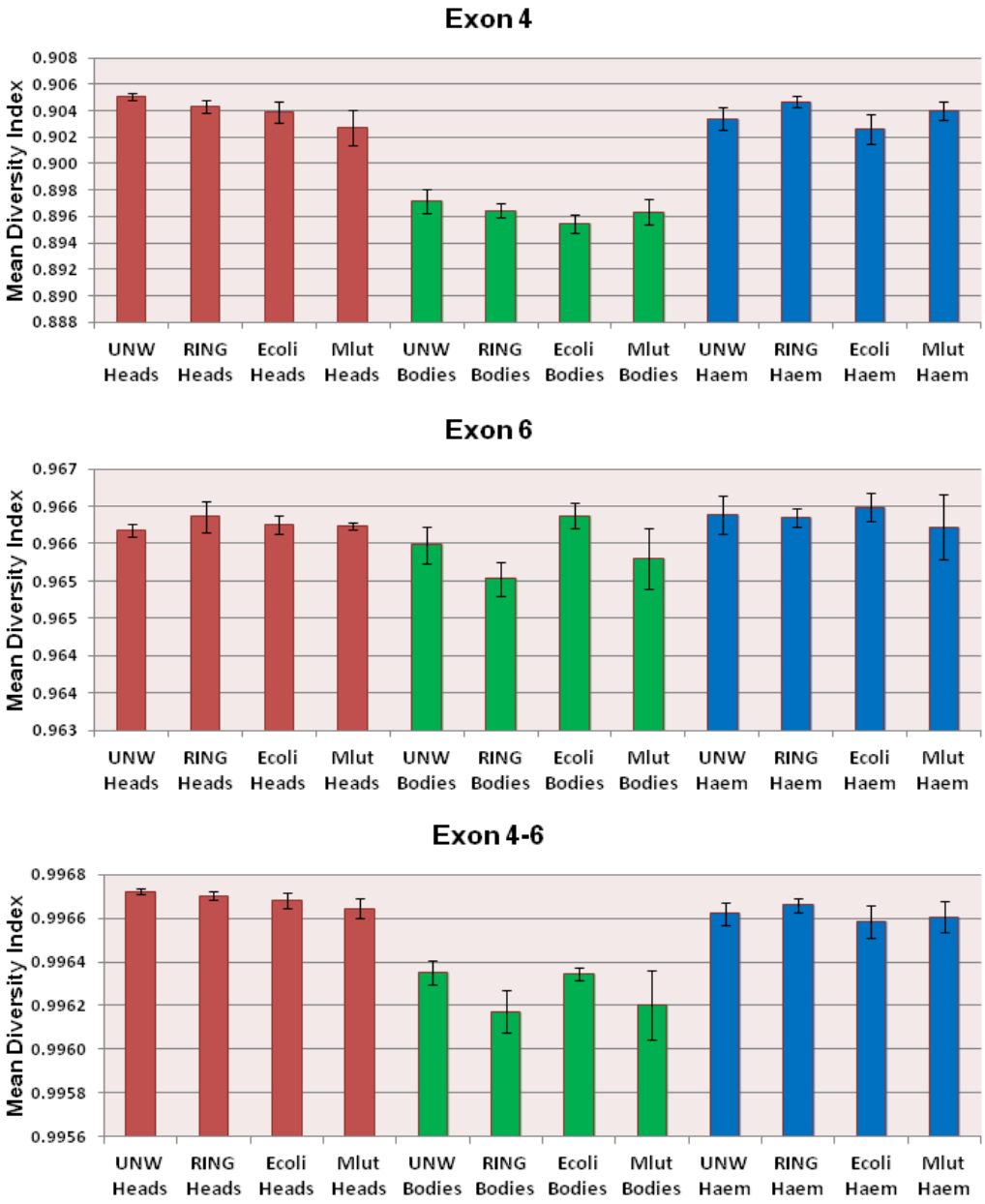


Figure 3: Expression diversity for exon 4, exon 6, and exon 4-6 combinations between tissues and treatments. I fitted a GLM to expression diversity data as: (1-D) ~ tissue + pathogen + tissue*pathogen. Exon 4, and exon 4-6 combination expression diversity was lower in bodies than any other tissue, and exon 6 expression diversity appeared to be similar across all tissues. No effects of pathogen or interaction were detected (see text). Diversity is represented as mean 1-D of six independent replicates. Error bars represent SEM calculated from six independent experiments. **Key**: UNW: unwounded, Ring: Ringers-wounded, Ecoli: *Escherichia coli*, Mlut: *Micrococcus luteus*, Haem: haemolymph.

Using MCMCglmm, I fitted models to the count data to test for an interaction between Dscam expression and experimental conditions. I firstly tested for effects of wounding. Table 1 outlines the analysis output which represents the variance structure. The model mixed well and the average effective sample size was over 7000 (see appendix, figure S10).

Table 1: Variance structure summarising a MCMCglmm model testing for effects of wounding on Dscam expression. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Primer-pair	0.001	0.0002	0.0001	[0, 0.002]	0.0
Replicate	0.001	0.001	0.00001	[0, 0.003]	0.0
Exon 4	**10.388**	**8.983**	**7.792**	**[3.27, 20.99]**	**53.0**
Exon 6	**7.479**	**7.248**	**6.869**	**[4.54, 10.74]**	**38.1**
Exon 4-6	**0.253**	**0.252**	**0.251**	**[0.21, 0.29]**	**1.3**
Exon 4: Tissue	**0.287**	**0.266**	**0.248**	**[0.11, 0.49]**	**1.5**
Exon 4: Wounding	0.007	0.004	0.0003	[0, 0.02]	0.0
Exon 6: Tissue	0.003	0.003	0.00004	[0, 0.007]	0.0
Exon 6: Wounding	0.001	0.0003	0.00001	[0, 0.002]	0.0
Exon 4: Tissue: Wounding	**0.035**	**0.033**	**0.031**	**[0.02, 0.06]**	**0.2**
Exon 6: Tissue: Wounding	0.002	0.002	0.00002	[0, 0.006]	0.0
Exon 4-6: Wounding	0.001	0.001	0.00001	[0, 0.004]	0.0
Exon 4-6: Tissue	**0.018**	**0.018**	**0.018**	**[0.007, 0.03]**	**0.1**
Exon 4-6: Tissue: Wounding	0.001	0.001	0.00002	[0, 0.005]	0.0
Residuals	1.136	1.136	1.132	[1.11, 1.16]	5.8

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

As expected following previous χ^2 analysis of constitutive expression, a considerable amount of the total variance explained by random effects in the model was attributable to exon 4, exon 6, and exon 4-6 combinations (Table 1). In addition, 1.5% of the total variance was explained by an exon 4-by-tissue effect, while smaller contributions to total variance were

attributable to an exon 4-by-tissue-by-wounding effect (0.2%) and an exon 4-6 combination-by-tissue effect (0.1%). The lower credibility bound of each of these effects did not reach zero. Otherwise, contributions to total variance were negligible for all terms in the model. Thus, only a weak effect of wounding was evident, but notable effects of tissue were detected.

I next tested for effects of pathogen. Table 2 outlines the analysis output representing the variance structure for the model. The model mixed well and the average effective sample size was above 7500 (see appendix, figure S11).

Table 2: Variance structure summarising a MCMCglmm model testing for effects of pathogen-exposure on Dscam expression. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Primer-pair	0.001	0.0004	0.0002	[0, 0.004]	0.0
Replicate	0.003	0.003	0.002	[0, 0.005]	0.0
Ex4	**10.861**	**9.254**	**6.977**	**[2.82, 22.31]**	**54.3**
Ex6	**7.234**	**7.022**	**6.488**	**[4.52, 10.36]**	**36.2**
Exon 4-6	**0.280**	**0.278**	**0.272**	**[0.24, 0.32]**	**1.4**
Ex4: Tissue	**0.351**	**0.329**	**0.287**	**[0.16, 0.59]**	**1.8**
Ex4: Pathogen	0.007	0.006	0.00004	[0, 0.02]	0.0
Ex6: Tissue	0.003	0.003	0.003	[0, 0.006]	0.0
Ex6: Pathogen	0.0003	0.0001	0.000002	[0, 0.001]	0.0
Ex4: Tissue: Pathogen	**0.021**	**0.021**	**0.019**	**[0.012, 0.032]**	**0.1**
Ex6: Tissue: Pathogen	0.0004	0.0002	0.00001	[0, 0.002]	0.0
Exon 4-6: Pathogen	0.0003	0.0002	0.000005	[0, 0.001]	0.0
Exon 4-6: Tissue	**0.018**	**0.018**	**0.019**	**[0.01, 0.03]**	**0.1**
Exon 4-6: Tissue: Pathogen	0.0004	0.0002	0.00001	[0, 0.002]	0.0
Residuals	1.227	1.227	1.227	[1.21, 1.25]	6.1

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

Here, nearly 2% of the total variance was explained by an exon 4-by-tissue effect, with smaller proportion of the total variance being attributable to an exon 4-6 combination-by-tissue effect (0.1%). In addition, a very small exon 4-by-tissue-by-pathogen effect was seen (0.1%). The lower credibility bound of each of these effects did not reach zero. Thus, effects of tissue were apparent, closely reflecting the results seen in the model testing for effects of wounding (Table 1), and only a weak effect of pathogen-exposure was evident.

To investigate whether I could detect differential expression of Dscam exon variants between different tissues and between different types of bacteria in *Drosophila*, I employed DESeq and MCMCglmm analyses to test the number of reads for each *Dscam* exon variant for being differential between treatments. Figure 4 shows heat maps representative of exon 4 and exon 6-expression for all treatments in the study. Strong patterns were seen for exon variants across all treatments for both exons 4 and 6, and patterns were detected between tissues. However, visual inspection suggested that no discernible patterns were obvious between treatments.

To firstly test for differential expression following wounding, I compared exon variant expression between unchallenged and sterile-wounded flies. Using DESeq, I found no difference in the expression of any exon 4 or exon 6 variant between the unwounded and Ringers-wounded flies in heads (exon 4: all q-values ≥ 0.90; exon 6: all q-values ≥ 0.98), bodies (exon 4: all q-values ≥ 0.82; exon 6: all q-values ≥ 0.74), or haemolymph (exon 4: all q-values ≥ 0.35; exon 6: all q-values ≥ 0.87). Following MCMCglmm analysis however, I found an exon 4.6-by-body-by-wounding effect ($p = 0.01$).

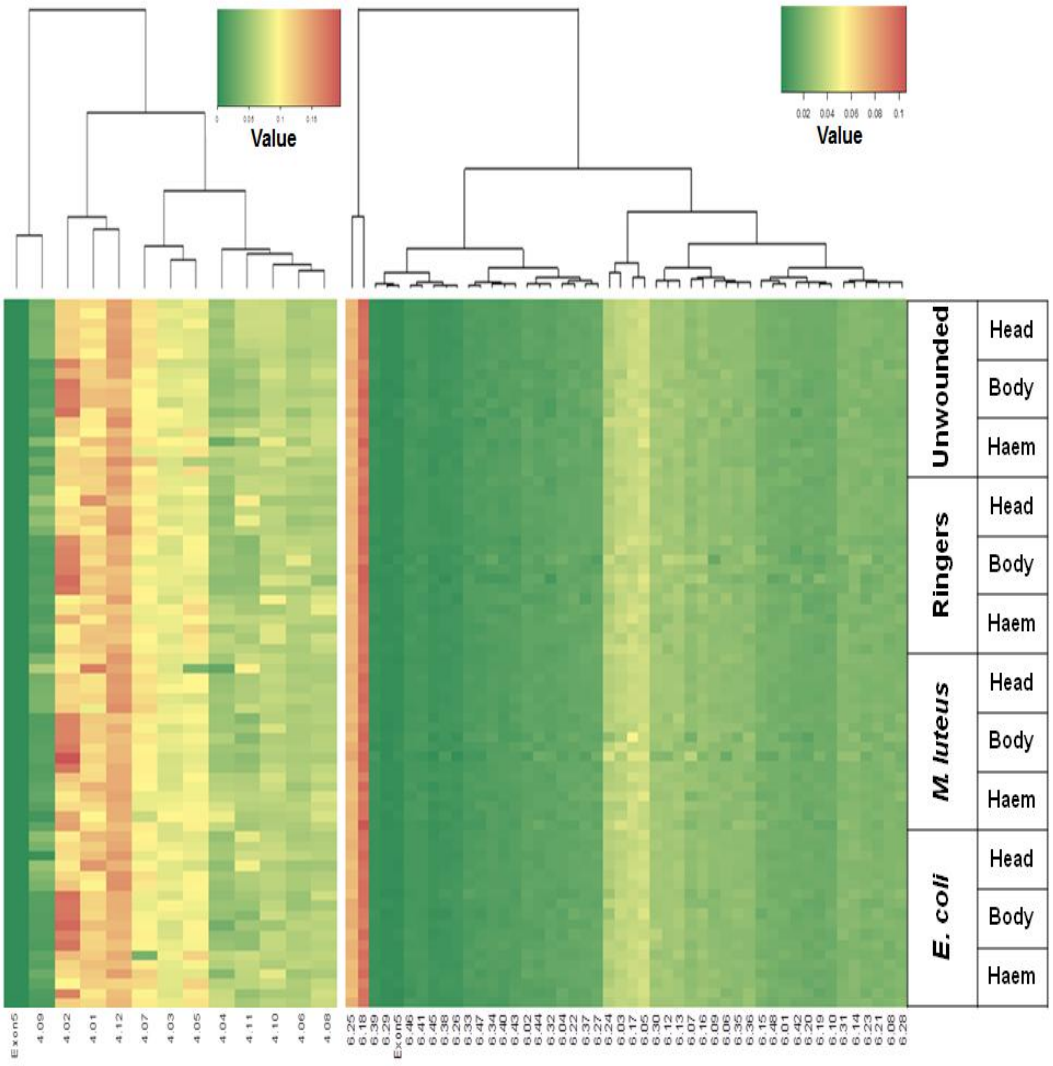


Figure 4: Heat map representation of relative expression of exon 4 (left block) and exon 6 (right block) variants for all treatments in this study. Patterns of expression for individual variants of both exons 4 and 6 across treatments and between tissues were evident. For example, exon 4.2 was clearly the most abundantly expressed exon 4 variant in bodies regardless of treatment, while exon 4.11 was expressed at lower levels in bodies than in heads or haemolymph in all exposures. Moreover, exon 4.9 appeared to be expressed at higher levels in heads than any other tissue. For exon 6, variants 6.18 and 6.25 were significantly more abundant than any other exon 6 variants across all treatments, while exons 6.29 and 6.39 were the least abundant across all treatments. No obvious patterns of expression were detected between treatments. Exon 6.11 was not detected and hence not included. The inclusion of exon 5 indicates relative abundance of exon-skipping for each exon cluster. Each tissue type, i.e. Head, Body, Haem (haemolymph) include 6 biological replicates, and each cell represents the proportional expression of a given exon variant in a single sample. Dendrograms result from hierarchical clustering of exon variant expression pattern using Euclidean distance measures.

To investigate differential expression between exon variants following septic-wounding, I used DESeq to compare expression in sterile-wounded flies with expression in flies exposed to the bacteria *E.coli* and *M. luteus*. I found no differences in expression of any exon 4 variants across any tissue type in response to *E. coli* (exon 4-expression in heads: all q-values ≥ 0.96; in bodies: all q-values ≥ 0.91; and in haemolymph: all q-values ≥ 0.92). Similarly, exposure to *M. luteus* revealed no significant differences in exon 4-expression (heads: all q-values ≥ 0.97; bodies: all q-values ≥ 0.98; haemolymph: all q-values ≥ 0.39). Moreover, exon 6 variant expression in any tissue was not affected by either *E. coli* or *M. luteus*. Following analysis with MCMCglmm, an effect on the expression of exon 4.4 was seen in haemolymph in response to *M. luteus* exposure ($p = 0.01$). The lower credibility bound did not reach zero for this effect, but was very close to zero. Thus, no detectable effects of pathogen-exposure were seen following analysis with DESeq, but a very small effect of *M. luteus*-exposure was seen on the expression of exon 4.4 in haemolymph following analysis with MCMCglmm.

I next tested if patterns of expression of exon 4-6 combinations were apparent between tissues and between treatments. Figure 5 shows a heat map of the expression of exon 4-6 combinations of which all represented an abundance of over 0.5% in at least one treatment in the study. As with individual exons, visual inspection suggested that patterns were evident across treatments. For example, transcripts containing the combinations 4.2-6-18 and 4.12-6-18 were the most abundantly expressed across all treatments which reflected the relatively high abundances of these variants independently (Fig. 5).

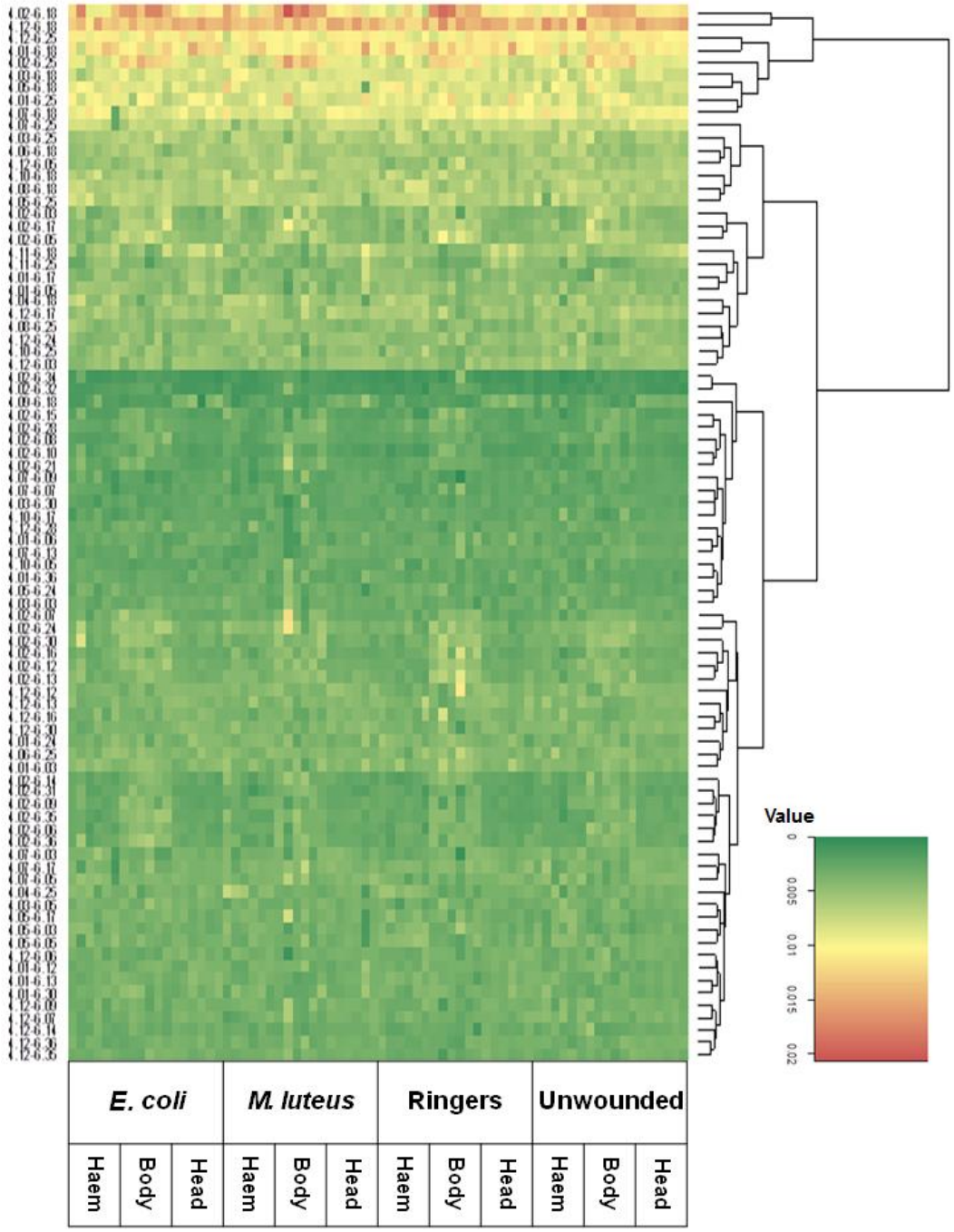


Figure 5: **Heat map representation of expression of exon 4-6 combinations of which each represented a minimum abundance of 0.5% in at least one treatment in this study.** Patterns of expression for exon 4-6 combinations across treatments were evident. For example, transcripts containing the combinations 4.2-6-18 and 4.12-6-18 were the most abundantly expressed across all treatments. No discernible patterns were obvious between treatments. Each cell represents the proportional expression of a given exon combination in a single sample. Dendrograms result from hierarchical clustering of exon combination expression pattern using Euclidean distance measures.

However, there was no difference between treatments. Applying DESeq analysis for the same pairwise comparisons carried out for individual exons (see above), I found no evidence of differential expression of any exon 4-6 combinations in response to wounding or pathogens in this study. Moreover, analysis using MCMCglmm detected no effects of wounding or pathogen-exposure on the expression of exon-exon combinations (see Tables 1 and 2). Thus, no significant changes of expression of exon 4-6 combinations between treatments were evident.

4.3.6. Correlating the tissue-specific data with those of Chapter 3

To test the consistency of the Illumina sequencing assay between two independent experiments, I tested for correlation between data obtained from whole flies (see Chapter 3, figure 3) with those obtained from each tissue in the present study (Fig. 1). Figure 6 shows a bar chart comparing proportional expression of exon 4 (top) and exon 6 (bottom) variants between the studies.

Data from Chapter 3 positively correlated with those of the present study. An especially strong positive correlation was found for exon 4-expression between whole flies and fly bodies (exon 4: whole flies with bodies, $r_s = 0.881$, p = <0.0001), but also the other tissues (exon 4: whole flies with heads, $r_s = 0.853$, p = 0.0004; exon 4: whole flies and haemolymph, $r_s = 0.874$, p = 0.0002). Strong positive correlation was also seen for exon 6 between the data sets (exon 6: whole flies with heads, $r_s = 0.789$, p = <0.0001; exon 6: whole flies and bodies, $r_s = 0.794$, p = <0.0001; exon 6: whole flies with haemolymph, $r_s = 0.784$, p = <0.0001). Thus, positive correlations were found between the data of two independent experiments that utilised Illumina-based sequencing technology to study *Dscam* gene variant expression in *Drosophila*.

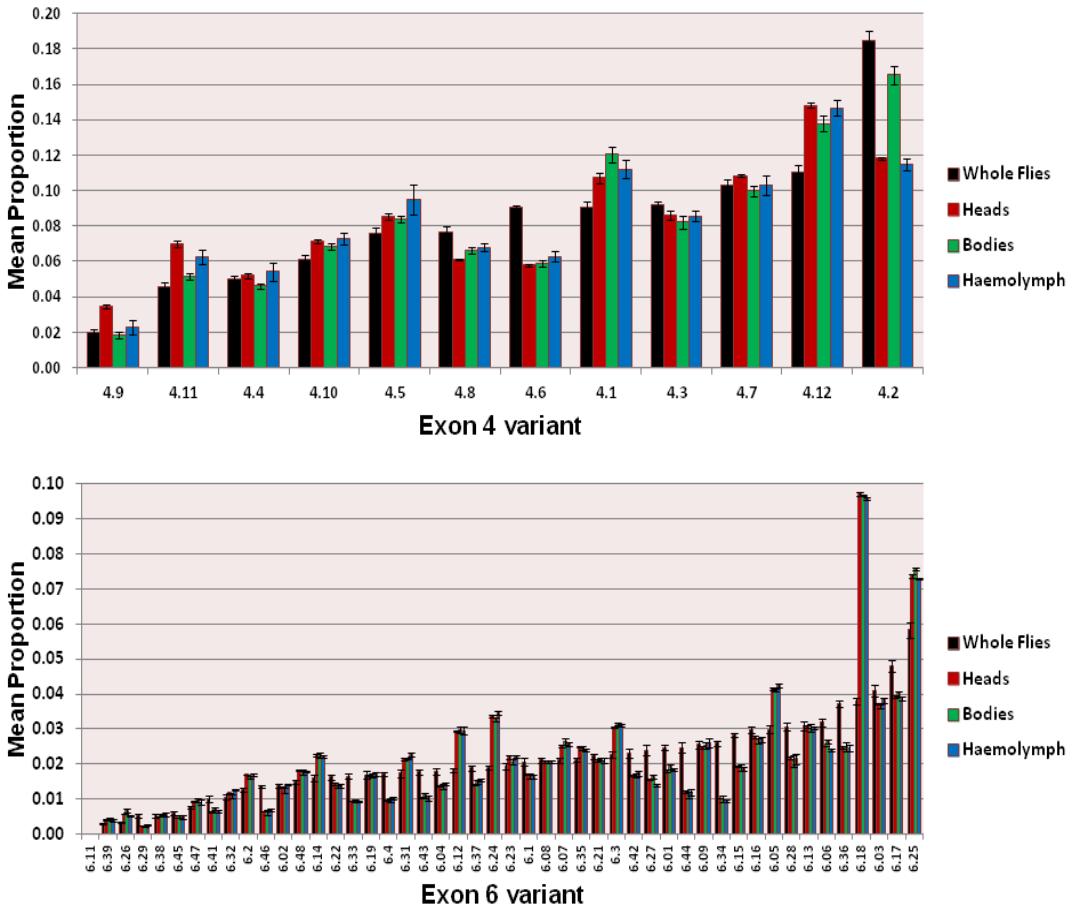


Figure 6: **Bar chart comparing Dscam expression data from whole flies of an independent study with those of tissues of the present study.** Expression data from both exons positively correlated between experiments (see text). Bars represent mean proportion of exon variant expression in pooled unwounded flies, and error bars represent SEM (whole flies, n = 10; each tissue, n = 6).

4.4. Discussion

4.4.1. Exon 4 and 6 variants are constitutively expressed in a non-random fashion in different tissues in Drosophila and exon 4, but not exon 6, is expressed in a tissue-specific manner

Exon 4 and 6 variants were expressed in a non-random fashion in all tissues assayed in this study. Moreover, exon 4 variants appeared to be spliced in a tissue-specific manner. For example, exon 4.2 was expressed in significantly higher amounts in bodies than in heads or

haemolymph (Fig. 1). Exon 4.2 was found to be the most abundantly expressed variant from the exon 4 cluster in whole flies in my previous experiment (see Chapter 3, figure 3), and was also the most abundant exon 4 variant found in both legs and wings in an independent SSCP assay in *D. melanogaster* (Celotto and Graveley 2001). As the body tissue in the present study contained legs and wings, the abundance of exon 4.2 in these appendages could be driving the pattern. Celotto and Graveley (2001) also discovered that exon 4.2 displayed the most striking developmental changes of any exon 4 variant. Only ~1% of transcripts in embryos contained this variant but the expression steadily increased throughout development becoming the most predominant exon 4 variant in adults, consistent with the appearance of legs and wings. The characteristics of exon 4.2-expression appears to be conserved in *D. yakuba* (Celotto and Graveley 2001), estimated to have diverged from *D. melanogaster* ~3-15 million years ago (Powell 1997; Li, Satta et al. 1999; Tamura, Subramanian et al. 2004; Obbard, Maclennan et al. 2012), indicating a relative importance of the tissue-specific function of exon 4.2 across different *Drosophila* species. Additionally, the expression diversity of exon 4 was shown to be lower in bodies than in heads or haemolymph. This indicates a less-restricted usage of exon 4 variants in heads and haemolymph compared to bodies, and may be driven largely by the relatively high abundance of exon 4.2 in body tissue of this study.

The expression of exon 4.9 was significantly higher in heads than in bodies or haemolymph (Fig. 1). In SSCP gel electrophoresis, exon 4.9 co-migrates with exons 4.5 and 4.7 (Celotto and Graveley 2001), therefore no comparison could be made between independent studies. Nevertheless, the data here suggest that exon 4.9 may have a more important role in heads/antennae than in any other tissues. Exon 4.11 was expressed significantly higher in heads than any other tissue. Around 7% of exon 4 variants in heads were exon 4.11 in this study and this reflected data from a past SSCP study (exon 4.11 in heads, $6.9 \pm \sigma$ 1.3%, see Celotto and Graveley 2001). Both experiments confirmed that the proportional inclusion of

exon 4.11 was significantly lower in the other tissues assayed (bodies and haemolymph in the present study, wings and legs in the SSCP experiment), and this suggests that exon 4.11 may also be a relatively more important variant in heads/antennae than other tissue types in *Drosophila*. Thus, the data confirm that choice of exon 4 variant is regulated as evidenced by the non-random expression of each variant, and also that alternative splicing of exon 4 variants happens in a tissue-dependent manner.

In contrast to exon 4-expression, exon 6 variant-expression revealed no evidence of tissue-specific alternative splicing. Nevertheless, exon 6 variants were expressed in a non-random fashion in all tissues indicative of some level of regulation (Fig. 1). Moreover, mean expression diversity of exon 6 did not differ between tissues, also indicating very similar non-random usage of exon 6 variants in all tissues (Fig. 3). To my knowledge this has never been reported. Watson et al (2005) using microarrays to compare exon 6-usage between haemocytes, fat bodies and brain tissue found that these variants were expressed in all three tissues, but the authors could not determine with any accuracy the relative abundances within each tissue. Thus, as in exon 4, non-random expression occurred in exon 6 in all tissues, but the proportional inclusion of exon 6 variants in transcripts found in all tissue types was similar confirming a lack of tissue-dependent expression for this exon.

Combinatorial splicing in unchallenged flies appears to be close to random in heads, bodies and haemolymph. However, separate tissues reflected the patterns seen in whole flies (Chapter 3) where most combinations were detected at expected frequencies but some appeared to be expressed non-randomly (Fig. 2). Indeed, MCMCglmm analysis detected small exon 4-6 combination-by-tissue effects in both models (see Tables 1 and 2), indicative of a small difference in exon 4-6 combination expression between tissues. Departures from randomness may be a result of PCR stochasticity. As discussed in depth in Chapter 3, effects of random PCR noise may have caused some exon-exon combinations to amplify more

efficiently than others, resulting in the patterns seen. However, it should be noted that no striking differences were found between individual replicates in this study. Nevertheless, it would be interesting to discover whether these same patterns of exon 4-6 combination expression occurred in future experiments in the same tissues under the same conditions.

As seen in Chapter 3 in whole flies, exon-skipping was detected in this study and in all tissues. Interestingly, despite the evaluation of Dscam transcript expression in very different tissue types, exon 6-skipping was detected at remarkably similar frequencies in each. Whether the phenomenon of exon-skipping is a result of splicing failure or is part of a regulatory mechanism, it does not appear to show any tissue-specificity for the exon 6 cluster at least.

4.4.2. Tissue-specific Dscam expression in response to wounding and pathogens

Dscam expression diversity was comparable between unwounded and sterile-wounded treatments, suggesting that *Drosophila* do not respond to tissue damage with profound changes in expression diversity of Dscam gene variants (Fig. 3). Moreover, I found no differential expression of any exon 4 or 6 variants in response to wounding using DESeq. However, I found a small exon 4-by-tissue-by-wounding effect following analysis with MCMCglmm (Table 1). Specifically, wounding appeared to affect the expression of exon 4.6 in bodies. However, unlike DESeq analysis, MCMCglmm does not correct for multiple testing. As such, care must be taken in interpreting the results.

E. coli and *M. luteus* are strong immune elicitors in *Drosophila* and provoke substantial upregulation of the immune markers Diptericin and Drosomycin, respectively (Lemaitre, Reichhart et al. 1997). Here I found comparable Dscam expression diversity between sterile-wounded and pathogen-exposed treatments, indicating overall expression diversity does not significantly change in response to pathogen-exposure in *Drosophila* (Fig. 3). This did not

reflect results seen in *Anopheles*, where Dscam expression diversity increased following exposure to *Plasmodium* parasites (see Chapter 2). A model used to examine the effects of pathogen-exposure detected a very small exon 4-by-tissue-by-pathogen effect. Specifically, analysis with MCMCglmm suggested that the expression of exon 4.4 in haemolymph was affected by *M. luteus*-exposure. However the effect did not dominate over tissue effects (Table 2). Lastly, following DESeq analysis of differential expression, I found no difference in expression between any exon 4 or 6 variants between Ringers and *E. coli* treatments, or between Ringers and *M. luteus* treatments, in any tissue type after adjusting for multiple testing. Thus, only weak evidence that *Drosophila* respond to the presence of pathogen-exposure with any marked changes of Dscam expression was found.

4.4.3. Supporting the sequencing assay as a reliable method for the study of alternative splicing of Dscam

As this Illumina-based sequencing method was relatively novel for the study of *Dscam* expression, it is important to determine as much as possible its reliability. As such, I looked for a correlation between the data of this experiment with those of a previous experiment (Chapter 3). I found positive correlations between the expression data from whole flies and those of the present study suggesting a high level of consistency between the independent experiments. Similarities included the expression of exon 4.2, which was the most abundantly expressed exon 4 variant in both studies, with exon 4.9 being the least abundant in both experiments. However, differences were found between the studies. For instance, the expression of some exon 6 variants was surprisingly different between the independent experiments despite the use of the same *Drosophila* strain. For example, in whole flies (Chapter 3), exon 6.25 was the most abundantly expressed variant from this exon cluster, while in the present study, exon 6.18 was the most abundant exon 6 variant in all tissues studied, with exon 6.25 the second most abundant.

The possibility of early-round PCR stochasticity here remains a possibility. However, this phenomenon would arguably be expected to have been more of a problem in the experiment reported in Chapter 3, which unlike the present study, did not benefit from pooling several replicate PCR products. Nevertheless, data from Chapter 3 correlated well with independent microarray data (see Chapter 3, Discussion), and data in the present study correlated well with Chapter 3. Furthermore, both the comparable relative expression of exon 6 across tissues, and the comparable frequency of exon 6-skipping-events across tissues, indicate a level of consistency in PCR amplification between samples in the present study. Thus, although early-round PCR stochasticity remains a threat with this method, it may not have significantly affected the landscape of Dscam gene variant expression seen in the fly. Therefore, the data from the present study may be a more accurate representation of Dscam gene expression as a result of pooling PCR products to reduce effects of PCR stochasticity. Regardless, future studies using this approach would no doubt benefit from complementary microarray and/or qRT-PCR assays of the relative abundances of exon variants to help control for any possible early-round PCR stochasticity in high-throughput sequencing assays of Dscam expression.

4.5. Conclusion

Previous research has shown exon 4-expression differs between tissues (Celotto and Graveley 2001) and this was detected here using Illumina sequencing. Exon 6-expression did not appear to be tissue-specific. However, neither expression of exon 4, exon 6, or exon 4-6 combinations appeared to be significantly affected by *Drosophila* immune elicitors. Nevertheless, the experiment could not rule out a small effect of challenge on Dscam expression in *Drosophila*. However, the data suggest that it does not dominate over the observed tissue-effects, where an exon 4-by-tissue effect represented nearly 2% of the total

variance explained by random effects in the models (Tables 1 and 2), while an effect of pathogen-exposure on exon 4 was a mere 0.1% (Table 2).

Nevertheless, at present two independent experiments in *Anopheles* (Chapter 2), and now two independent experiments in *Drosophila* (Chapters 3 and 4) all suggest the presence of a small but detectable exon 4-by-challenge effect. Future Illumina-based sequencing of Dscam expression in *Drosophila* may be best served by examining the technology against a low background-noise model. For instance, well-characterised pathogen-challenge of *Drosophila* S2 cells, cultured immune-competent cells thought to have derived from macrophage-like lineage of haemocytes (Schneider 1972), would exclude background noise of non-immune-related tissues and hopefully direct further informative studies of the specifics of Dscam gene expression in invertebrate immunity.

Chapter 5: Tissue-specific Dscam gene expression in response to within-pathogen species diversity in the crustacean *Daphnia magna*

This chapter was written with comments from Darren Obbard and Tom Little. Statistical analyses were done in collaboration with Darren Obbard. The quality trimming, counting and sorting of raw Illumina sequencing reads was done using a custom R pipeline written by Darren Obbard. Illumina library construction, sequencing and base-calling were performed by the GenePool Sequencing Facility (University of Edinburgh). *Daphnia* rearing, immune challenge, and tissue separation was carried out by Carolyn Riddell and Phil Wilson. RNA isolation was carried out by Carolyn Riddell.

5.1. Introduction

Down syndrome cell adhesion molecule, Dscam, was first discovered and described as a single-pass neuronal membrane receptor of the vertebrate nervous system (Yamakawa, Huo et al. 1998). Soon after, a hyper-diverse form of the gene was discovered in *Drosophila*, where it functions as an axon guidance receptor in invertebrate neuronal development (Schmucker, Clemens et al. 2000). A capacity to produce many thousands of isoforms from a single gene through the mechanism of alternative splicing is an essential feature of its role in controlling nerve cell interactions in invertebrates (Chen, Kondo et al. 2006; Hattori, Demir et al. 2007). So far, this hypervariable form of the receptor has not been discovered outside the Pancrustacea, a clade consisting of the subphyla Hexapoda (which includes the insects), and Crustacea. Several studies have predicted that the alternative splicing *Dscam* gene originated along the lineage leading to the insects and crustaceans, which are thought to have diverged from a common ancestor around 450-500 million years ago (Crayton, Powell et al. 2006; Brites, McTaggart et al. 2008; Lee, Kim et al. 2010; Armitage, Freiburg et al. 2012).

Both insect and crustacean Dscams have been implicated in immunity. In *Anopheles* and *Drosophila*, depletion of Dscam in immune cells results in a significant reduction of phagocytic activity (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006). In addition, pathogen-specific Dscam isoforms have been reported to be expressed in the *Anopheles* mosquito in response to diverse pathogens (Dong, Taylor et al. 2006), and in the crustacean, *Pacifastacus leniusculus*, in response to bacteria-exposure (Watthanasurorot, Jiravanichpaisal et al. 2011). In another crustacean, *Daphnia magna*, diverse Dscam isoforms have been detected in haemocytes, the effector cells of immunity, suggesting a role in the immune response as reported in the insects (Brites, McTaggart et al. 2008).

Interestingly, *Daphnia* Dscam appears to have fundamental differences compared to the insect Dscams. For example, whereas two different transmembrane domains have been discovered in the Dscams of insects, *Daphnia* Dscam has only one (Brites, McTaggart et al. 2008). Perhaps surprisingly, Dscam of the crayfish, *P. leniusculus*, shares more similarity with the insect than *Daphnia* by having two transmembrane domains (Watthanasurorot, Jiravanichpaisal et al. 2011). This could imply a more restricted role for *Daphnia* Dscam compared with Dscams in other members of the Pancrustacea. However, multiple cytoplasmic domains have been found in both *Daphnia* and *Pacifastacus* Dscams, but not in those of the insects (Brites, McTaggart et al. 2008; Watthanasurorot, Jiravanichpaisal et al. 2011). Thus, Dscams in crustaceans may have different roles to play in both neuronal wiring and immunity compared with other Dscams found across the clade.

The immunoglobulin (Ig) and fibronectin type III (FnIII) domain structure of *Daphnia* Dscam is otherwise the same as the insect Dscam. The arrangement is 9Ig-4FnIII-Ig-2FnIII, and as in insects, the three highly variable exon clusters of *Daphnia* encode the first half of Ig2 and Ig3, and the entire Ig7. Figure 1 illustrates the protein and gene structure of *Daphnia* Dscam compared with *Drosophila* Dscam. Like Dscam in many *Drosophila* species, different *Daphnia* species have very similar copy numbers at each variable exon cluster. For instance, in *D. magna*, variable exons 4, 6 and 11 contain 8, 24 and 17 alternative variants, respectively, while in *D. pulex*, the same variable exons contain 8, 26 and 16 variants, respectively (Fig. 1). Including four alternative cytoplasmic tail forms, *D. magna* Dscam has the capacity to produce up to 13,056 isoforms through alternative splicing, and *D. pulex* can

produce up to 13,312 isoforms.

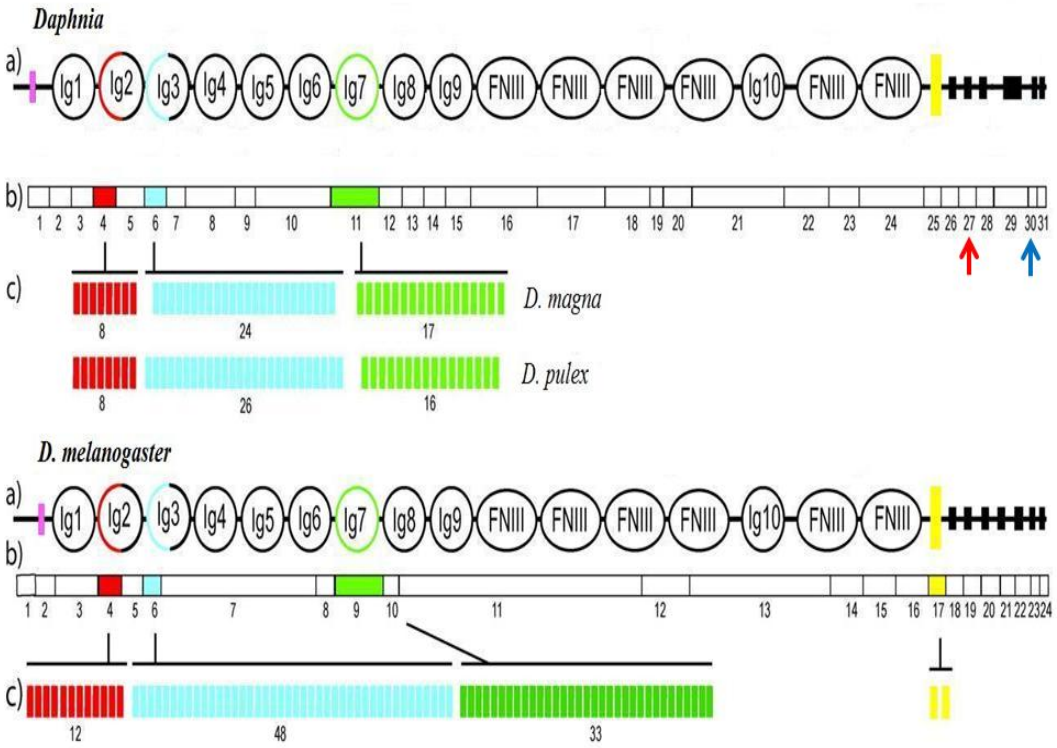


Figure 1: Protein and gene structure of *Dscam* in *Daphnia magna*, *D. pulex*, and *Drosophila melanogaster*. (a) Protein structures of *Daphnia* and *Drosophila* Dscams, (b) Gene structures of *Daphnia* and *Drosophila* Dscams, and (c) Variable exon clusters of *Daphnia* and *Drosophila* Dscams. Crustacean protein structure is comparable to insect protein structure with a 9Ig-4FnIII-Ig-2FnIII arrangement. Unlike insect Dscams which have two variants for the transmembrane region (shown as vertical yellow lines), *Daphnia* Dscam has only one. Moreover, *Daphnia* Dscam can express 4 different forms of its cytoplasmic tail. The cytoplasmic tail is coded by exons 26 to 31, however exon 30 (blue arrow) can be skipped, resulting in exon 31 being translated in an alternative reading frame. Additionally, exon 27 (red arrow) can also be skipped, facilitating two further possibilities. Figure modified from (Brites, McTaggart et al. 2008).

As *D. magna* and *D. pulex*, the only studied *Daphnia* species, are thought to have diverged ~7-15 mya (Haag, McTaggart et al. 2009), and the variable exons are believed to have diverged before the split of these two species (Brites, McTaggart et al. 2008), the exons

appear to be under similar selective constraints. Similarities between insect and crustacean Dscams suggest that the receptor may share comparable functions across the Pancrustacea, however that *Daphnia* and other crustacean Dscams have multiple intracellular signalling domains suggests that outcomes to receptor activation may be more diverse in these species.

Daphnia are planktonic freshwater crustaceans of the order Cladocera, and are thought to be keystone species in freshwater pelagic habitats such as ponds and rivers (Steiner 2002). They are filter feeders of small suspended organisms such as green algae, and are frequently infected with various parasites including microsporidians, fungi and bacteria (Green 1974; Ebert, Payne et al. 1997; Stirnadel and Ebert 1997). Female *Daphnia* reach maturity at around 10 days and produce clutches of parthenogenetic eggs after each adult molt (Ebert 2005). They reproduce by cyclical parthenogenesis, where they can switch between apomictic parthenogenesis and sexual reproduction. Sexual reproduction is usually triggered by environmental cues, and in the absence of these cues *Daphnia* can be kept in the laboratory in a state of clonal reproduction (Carvalho and Hughes 1983; Hobaek and Larsson 1990; Slusarczyk, Dawidowicz et al. 2005). Importantly, clonal reproduction of *Daphnia* facilitates not only the study of genetics but also of host-parasite interactions (Ebert, Zschokke-Rohringer et al. 1998). In the present study, I used a single clone of *Daphnia magna* to investigate Dscam expression in response to parasite-exposure.

The obligate endoparasitic gram-negative bacterium, *Pasteuria ramosa*, has been found to frequently infect *D. magna* in the wild and has a very high virulence (Ebert 2008). It is an endospore-forming bacterium transmitted horizontally through the ingestion of waterborne spores only (newborns are always free of infection) (Ebert, Rainey et al. 1996). The fitness consequences of infection are severe (reproduction is halted), and infection can cause death after around 30 to 50 days post-infection (Ebert, Zschokke-Rohringer et al. 1998). *P. ramosa* also causes parasite-induced host-gigantism; characterised by an observable increase in body

size compared to that of an uninfected individual, a phenomenon believed to be adaptive for the parasite (Ebert 2005). Importantly, *P. ramosa* shows a high degree of genotype-specificity within species. It has been shown that host-clones differ in their resistance and susceptibility to different isolates of *P. ramosa* (Carius, Little et al. 2001).

Previous research has found that a cellular immune response occurs in *D. magna* following exposure to *P. ramosa* (Auld, Scholefield et al. 2010). Specifically, haemocyte counts in susceptible hosts increased significantly after exposure to the parasite, while in resistant hosts, haemocyte counts remained at control levels. Accordingly, the expression of high Dscam diversity detected in *Daphnia* haemocytes (Brites, McTaggart et al. 2008), together with the implication of Dscam's involvement in phagocytosis (carried out by haemocytes) in other species (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011), make this model suitable for the study of Dscam expression in response to infection. Moreover, as *Daphnia* reproduces mostly clonally, this permitted the study of Dscam expression in a controlled genetic background, excluding the confounding effects of germline polymorphisms introduced by sexual reproduction.

Here, *D. magna* was exposed to two different strains of *P. ramosa*. One of which the host-clone is known to be highly susceptible to infection, and the other highly resistant to infection (Carius, Little et al. 2001). I hypothesised that if Dscam is involved in the immune response of *Daphnia* to a natural parasite, the Dscam expression profile may reflect differences between these exposures, especially in a host-parasite association known to show strong host-genotype-parasite-genotype interactions (Carius, Little et al. 2001; Ebert 2008). I examined Dscam gene expression in response to parasite-exposure in three different tissue types. Haemolymph was chosen as it contains haemocytes, immune cells of the host. Expression was also examined in guts. The gut is the point of entry for *P. ramosa* infection

of *Daphnia* (Duneau, Luijckx et al. 2011) and thus an appropriate tissue to examine Dscam expression. Furthermore, this interaction represents an early stage of the infection process. Finally, the remainder of the host (hereafter, 'carcass') was collected and also examined for effects of parasite-exposure. Exon 4, exon 6, and exon 4-6 combination-usage was detected by high-throughput sequencing (see Chapters 3 and 4) and counts were assigned to each exon variant or exon-exon combination for each treatment. I aimed to (1) detect the patterns of constitutive expression in different tissue types in *D. magna* not exposed to parasites, and (2) determine whether infection-responsive alternative splicing of *Daphnia* Dscam occurs in response to different parasites and in different tissue types.

5.2. Methods

5.2.1. Daphnia rearing

*Daphni*a were cloned in the lab by propagating iso-female lines under constant environmental conditions. The *Daphnia* clone used in this study was originally isolated from a pond near Gaarzerfeld in Northern Germany and has been kept in the laboratory as a clonal line for many years (Carius, Little et al. 2001). Experimental animals were isolated from the second clutch after three asexual generations were allowed to pass to minimise conditional variation and to increase independence of replicates. Environmental effects were kept constant for all replicates by using uniform conditions throughout. *Daphnia* were kept in 250ml jars of synthetic pond water (sodium bicarbonate ($NaHCO_3$), calcium chloride ($CaCl_2.2H_2O$ dihydrate), selenium oxide (SeO_3), and Instant Ocean® sea salt concentration), replaced every two days during rearing and throughout the duration of the experiment. Animals were fed $5x10^6$ chemostat-grown *Chlorella vulgaris* (non-motile colonial single-cell green algae) per day. Experimental animals were all female, and aged for 5 days at 20°C with a 16:8h light:dark cycle prior to parasite-exposure.

5.2.2. Immune challenge

Spores of the obligate endoparasitic bacterium *P. ramosa* were originally obtained from the grinding of cadavers of deceased wild-caught *D. magna* females and maintained by passage through susceptible female *Daphnia* as they cannot be cultured *in vitro*. As such, the spore solution used in this experiment was made by homogenising infected hosts in double distilled H_20 (ddH_20). Two different strains were used, a strain to which the host is known to be susceptible to (*s*), and another to which the host is known to be resistant to (*r*) (Carius, Little et al. 2001). Uninfected *D. magna* were homogenised in ddH_20 and used as a parasite-free control exposure.

Immune challenge was carried out in 96-well plates. Dosage was 5×10^4 parasite spores per *Daphnia* for a 4-hour exposure time (previous research has shown a cellular immune response at 4 hours; see Auld, Scholefield et al. 2010). After 4 hours, hemolymph was withdrawn from each individual in cell extraction chambers containing ice-cold anticoagulant buffer (see Lavine, Chen et al. 2005 for details) by piercing the heart with a 25-gauge needle and allowing it to drain by capillary action directly into RNAlater storage solution (Ambion). The hemolymph and RNAlater were then mixed thoroughly by pipetting, and removed to -80°C. Guts were then carefully dissected, leaving the rest of the carcass, and these tissues were immediately and separately placed into RNAlater and removed to -80°C. The *Daphnia* were exposed to one of three treatments (control, spore *s*, or spore *r*), and the experiment was replicated 8 times. Every treatment contained 6 *Daphnia*, of which one was allowed to grow for up to 30 days to indicate infection status of the replicate. The other 5 individuals were used for the tissue-extraction process. A total of 144 *Daphnia* were used in this experiment.

5.2.3. RNA isolation, reverse transcription, PCR and Illumina Sequencing

RNA was extracted from *Daphni*a tissues using a standard TRIzol extraction protocol (Invitrogen). Total RNA was treated with RNase-Free DNase I (Ambion) to remove genomic DNA contamination, and complimentary DNA was synthesised using a Promega Reverse Transcription system with random hexamers. Reverse transcription conditions were as Chapters 3 and 4. Primers were designed to amplify a fragment of *D. magna Dscam* spanning from exon 3 to exon 7 (primer sequences were: (F) 5' – GACATCCACGTCCGAGCAG – 3', (R) 5' – CTTCCTTCCACGTATTTGTACCAT – 3'). Thus, PCR amplicons comprised variable exons 4 and 6, and the conserved exon 5. All forward primers incorporated unique molecular ID (MID) tags on the 5' end to assign reads to their treatments in multiplexed sequencing reactions. PCR conditions were as Chapters 3 and 4. PCR was repeated 5 times for each treatment and products were pooled to reduce effects of early-round PCR stochasticity. PCR products were quantified using the Qubit® 2.0 Fluorometer (Invitrogen), and all MID-tagged samples were mixed evenly based on quantification. The final mixed product was concentrated using ethanol precipitation, and sequenced in an Illumina Genome Analyzer II (service provided by the GenePool Sequencing Facility, University of Edinburgh).

5.2.4. Data analysis

Raw data handling

Raw data was handled exactly as Chapters 3 and 4. Sequencing data was pre-processed to produce better alignment results. Bad quality ends were cut, and poor quality sequence was trimmed by setting stringency parameters as Chapters 3 and 4. Minimum read length was set at 60bp and reads without a valid MID-tag were discarded.

Statistical Analyses

Analysis of constitutive exon expression within different tissues: I tested for a departure from even expression in different tissue types in *D. magna*. Pearson's chi-square (χ^2) tests were carried out on the raw counts of exons 4 and 6 of control *D. magna* to determine whether the frequency distribution of each exon variant was consistent with a random expression model (i.e. all exon variants are equally likely to be expressed; expectation calculated as 1/n variants at a given exon cluster). I tested exon 4-6 combinations in control *D. magna* for a departure from even expression using Fisher's Exact Tests on raw count data for each treatment due to low counts for some combinations. Generalised linear mixed models using MCMCglmm (Hadfield 2009) were also implemented to reveal non-random constitutive expression in different tissues by providing estimates of the proportion of variance explained by exons 4 and 6. See Chapter 3 Methods for description of MCMCglmm package. All analyses were handled in R (R Development Core Team 2011).

Comparisons of constitutive exon expression between different tissues: I compared the constitutive expression of exons 4 and 6 between tissues to test for tissue-specific Dscam expression using DESeq and MCMCglmm. I used DESeq (Anders and Huber 2010) to test for differential expression of exons 4 and 6 between tissues. DESeq is widely used to analyze count data associated with different genes. Here, gene variants were treated as equivalent to genes for the purposes of analysis. DESeq corrects for multiple testing using Benjamini-Hochberg's FDR (Benjamini and Hochberg 1995), and here FDR was controlled at 5% ($q = 0.05$). See Chapter 3 Methods for description of DESeq package. I used MCMCglmm to infer tissue-specific expression by providing estimates of the proportion of variance explained by exon-by-tissue effects in a generalized linear mixed model. All analyses were handled in R (R Development Core Team 2011).

Analysis of infection-responsive Dscam splicing in D. magna: I used three approaches to detect immune-responsive Dscam splicing. I estimated expression diversity using Simpson's index of diversity (Simpson 1949), and I fitted a general linear model (GLM) to indices of diversity to test for effects of tissue, pathogen, and an interaction between both as: Simpson's Index of Diversity (1-D) ~ tissue + parasite + tissue*parasite. I then used two complementary analyses to detect differential expression in response to parasites. Firstly, I used MCMCglmm to infer the effect of parasite-exposure on Dscam expression patterns. Specifically, I tested Dscam expression between control and parasite-exposed *Daphnia* (i.e. to investigate any effects of different parasite exposures and in different tissues) by fitting the model: reads ~ tissue + parasite + replicate + primer-pair + exon 4 + exon 6 + exon 4-6 + exon 4: tissue + exon 4: parasite + exon 6: tissue + exon 6: parasite + exon 4: tissue: parasite + exon 6: tissue: parasite + exon 4-6: tissue + exon 4-6: parasite + exon 4-6: tissue: parasite. I tested the fixed effects of tissue and challenge, and all interaction terms (outlined in Results). Relative expression was transformed as cube-root of percentages (expressed as reads per million) to ensure normally distributed residuals (see appendix figure S12 for Q-Q plot of distribution before and after transformation). Two runs of 1,000,000 MCMC iterations were completed for each model. Each run was sampled every 100 steps and the first 10% of iterations was discarded as burn-in. Secondly, I used the DESeq R package to detect differentially expressed variants in response to parasite-exposure. Specifically, I tested for differential expression of exons 4 and 6, and combinations thereof, between control and parasite strain *s*- or parasite strain *r*-exposed *Daphnia*. FDR was controlled at 5% (q = 0.05). All analyses were handled in R (R Development Core Team 2011). After assessment of results MCMCglmm analysis was repeated excluding data from haemolymph treatments (see appendix figure S13 for Q-Q plot of distribution before and after transformation with haemolymph treatments excluded).

5.3. Results

5.3.1. Illumina sequencing assay output

Data received from the sequencing facility had 134,498,169 read pairs. After quality trimming I had 19,719,327 reads which were still available in pairs of reads at least 60nt long. Sequences without a valid MID-tag or affected by mis-priming or template-switching were then discarded and a total of 14,265,039 million read pairs remained and were attributable to exons 4 and 6, and combinations thereof. The mean number of high quality reads from each PCR primer-pair was 200,915 (sequencing depth ranged from 44,787 up to 1,907,737).

5.3.2. Detecting differences in constitutive exon variant expression in different tissues

To test for uneven expression of exons 4 and exon 6 variants in different tissues in *D. magna* not exposed to parasites, I used Pearson's chi-square and MCMCglmm analyses. I found significant differences in relative expression between variants of both exons 4 and 6 in all three tissue types (Exon 4 gut: $\chi^2 = 27676.5$, DF = 7, p = <0.0001; Exon 4 carcass: $\chi^2 = 51853.8$, DF = 7, p = <0.0001; Exon 4 haemolymph: $\chi^2 = 68805.3$, DF = 7, p = <0.0001; Exon 6 gut: $\chi^2 = 12577.6$, DF = 23, p = <0.0001; Exon 6 carcass: $\chi^2 = 22828.3$, DF = 23, p = <0.0001; Exon 6 haemolymph: $\chi^2 = 133995.0$, DF = 23, p = <0.0001) (Fig. 2). Using MCMCglmm, I found significant effects of exon 4 and exon 6-expression, indicating non-random expression of both exons (see Tables 1 and 2). Thus, variants of exons 4 and 6 were expressed in a non-random fashion in different tissues in *D. magna*.

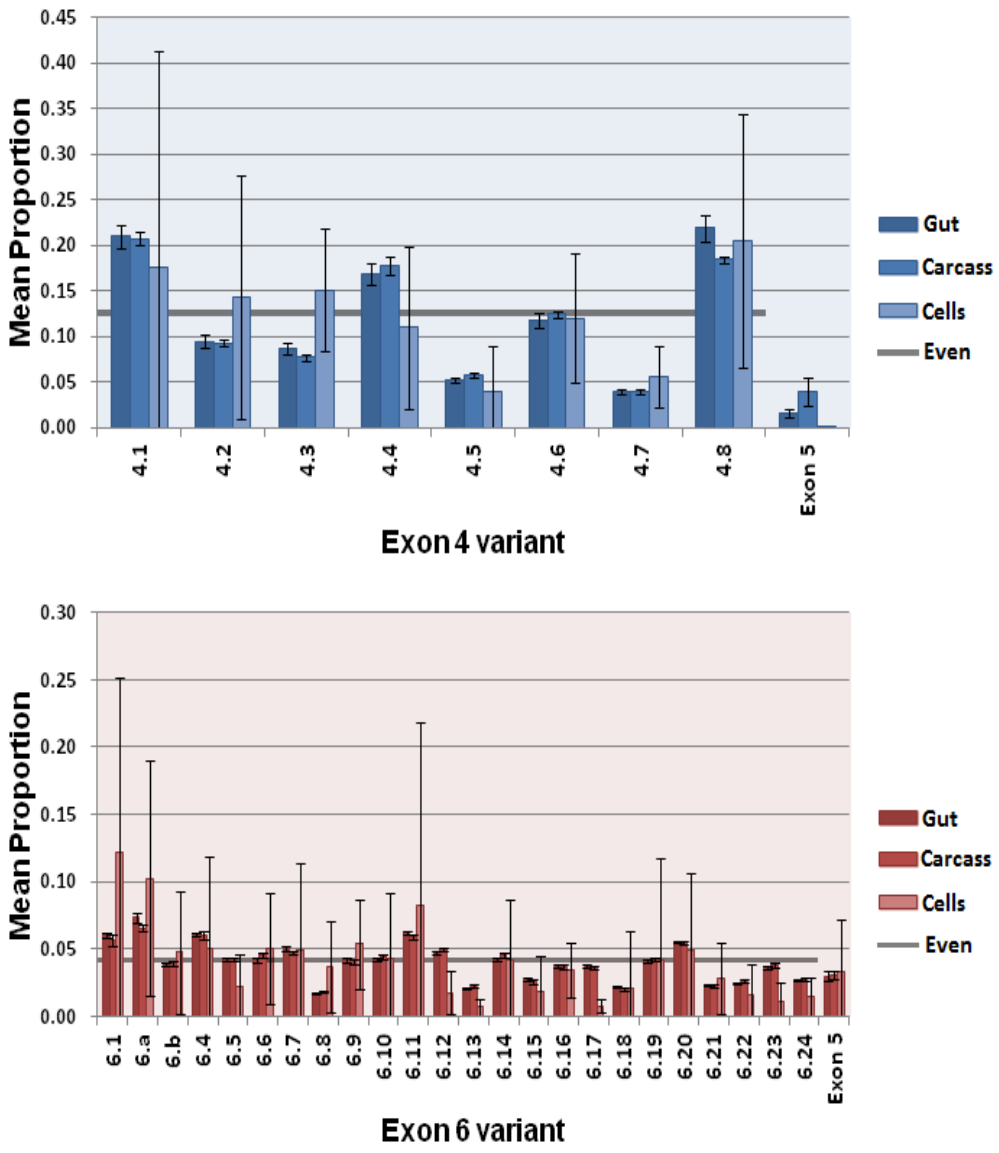


Figure 2: Bar chart representing relative constitutive expression of each exon 4 and 6 variant in different tissues in control *D. magna*. Exons 4 and 6 in guts, carcasses and haemolymph (cells) were expressed in a non-random fashion (see text). The inclusion of exon 5 indicates the mean proportion of exon-skipping for each exon. Even expression is shown as a horizontal line and based on a random expected proportion of all variants equally (i.e. even expression = 1/number of variants at a given exon cluster). Error bars represent standard deviation calculated from eight independent experiments.

5.3.3. Constitutive exon combinatorial splicing in different tissues of D. magna

I next compared the constitutive combinatorial splicing between the tissue types of *D. magna* and attempted to determine whether splicing was random or non-random in each tissue in the absence of parasites. Figure 3 shows a heat map of relative expression levels of all possible exon 4-6 combinations in each tissue and visual inspection suggests that the splicing of these two exons is close to random (i.e. conditional on the frequency of each exon 4 and 6 variant independently) in carcasses and guts, but that some exon-exon combinations were detected at frequencies deviating from random expression in these tissues. Indeed, Fisher's Exact Tests carried out on each exon 4-6 combination confirmed while most exon-exon combinations were evenly expressed, some combinations appeared to be non-randomly expressed. In contrast to carcasses and guts, much of the splicing detected in haemolymph deviated from random expression. Notably, I found that exon combination-expression in haemolymph was highly variable between individual replicates, whereas exon-exon splicing in guts and carcasses was similar between the biological replicates.

Overall, all eight expected exon 4 variants and an expected total of 24 exon 6 variants were expressed in all treatments in this study. However, exons 6.2 and 6.3 could not be confirmed with reference to the genome, and two alternative alleles not represented in the reference genome were detected and may be a result of allelic variation of these exons. The unidentifiable sequences were designated 6.aa and 6.bb for the purposes of further analysis. I aligned the unidentifiable sequences with the known exon 6 sequences and I constructed a phylogenetic tree using maximum likelihood methods for DNA sequences (Fig. 4). I found sequence 6.bb to be most similar to the *D. magna* exon 6.3 variant, while 6.aa was most similar to exon 6.17 (Fig. 4). In contrast to exon 6, no allelic variations of exon 4 were detected.

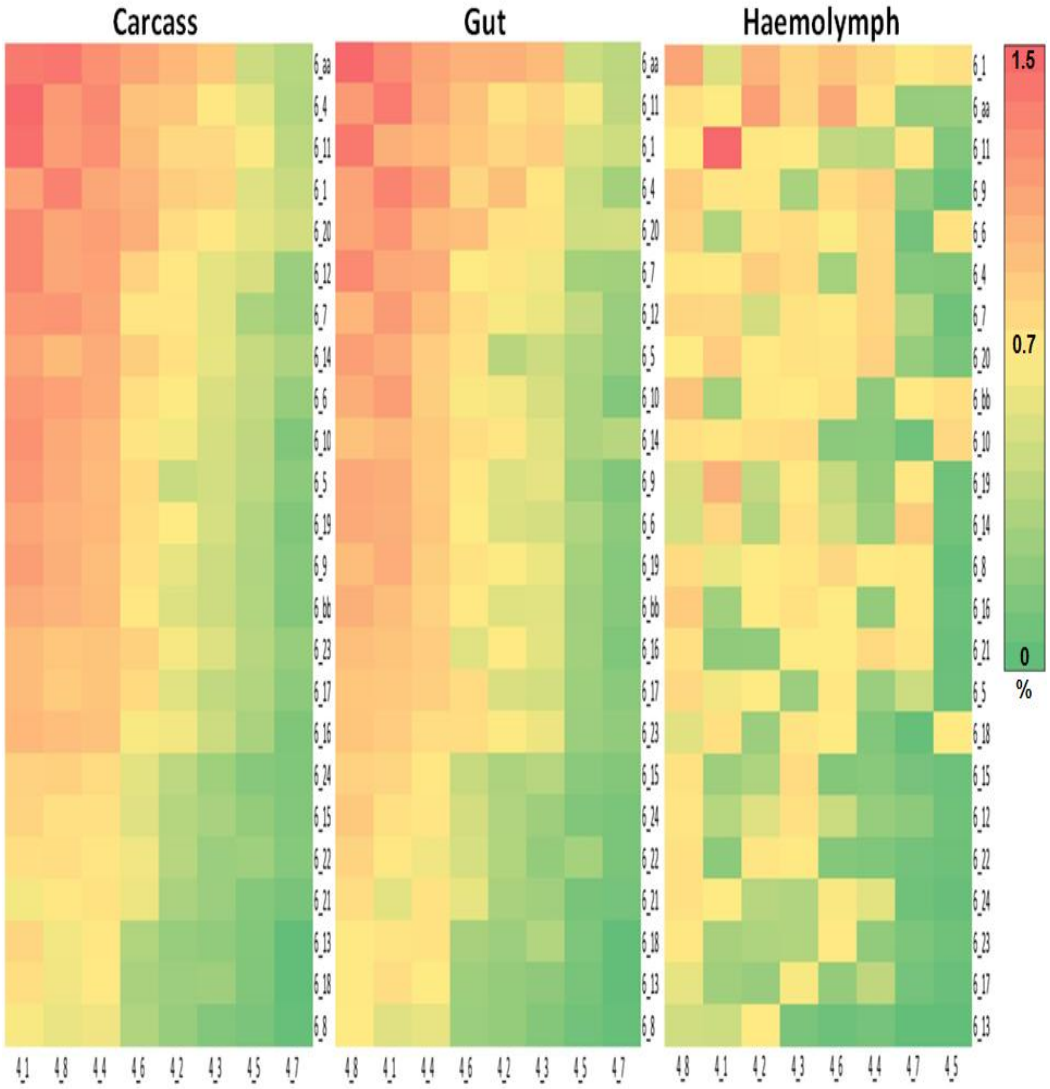


Figure 3: Heat map representation of the relative expression of all possible exon 4-6 combinations in carcass (left block), gut (middle block), and haemolymph (right block) in *D. magna* not exposed to parasites. Combinatorial splicing appeared to be close to random in gut and carcass tissues, but some combinations were detected at non-random frequencies. Splicing in haemolymph showed more non-random association than the other tissues. Each cell represents the proportion of an exon 4-6 combination, calculated as the mean of biological replicates of control treatments for each tissue (n=8).

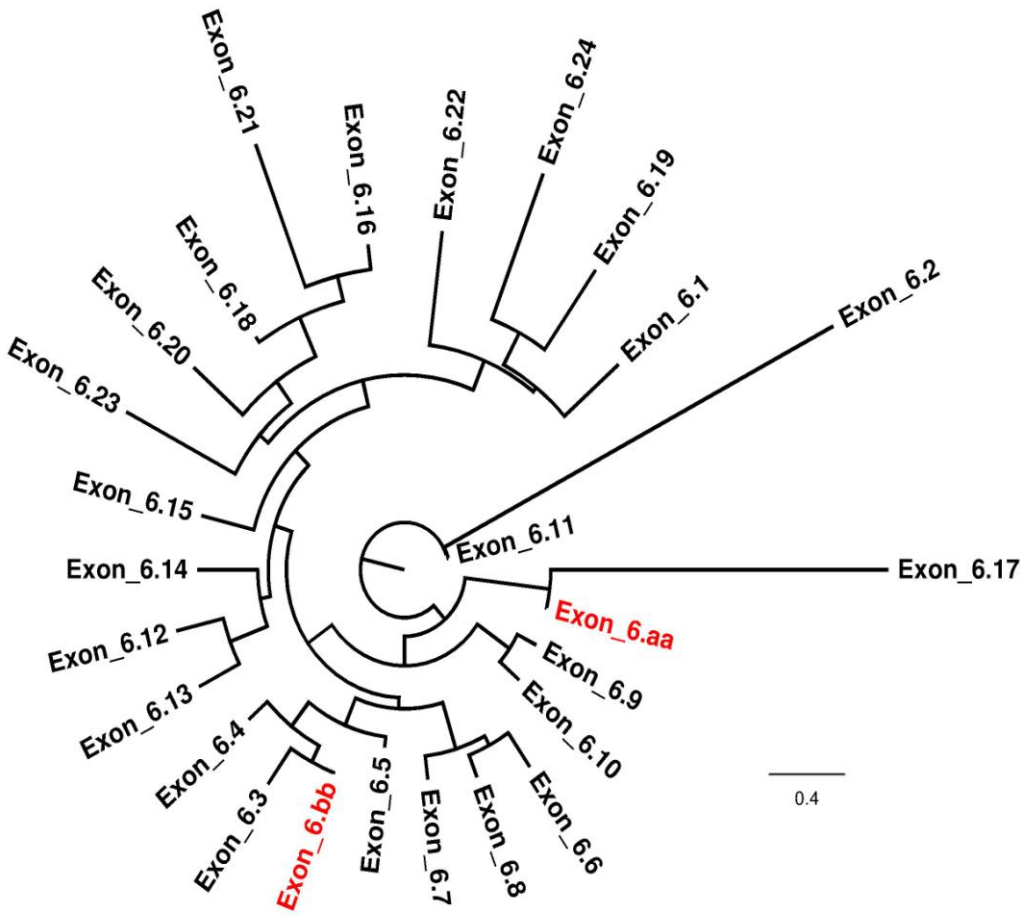


Figure 4: Phylogenetic relationships between *D. magna* exon 6 variants from the reference genome and two allelic variations found in this study. The maximum likelihood method for DNA sequences was used to construct the tree. Variant 6.bb was closest to 6.3, while variant 6.aa was closest to exon 6.17. Sequences were aligned using BioEdit Sequence Alignment Editor v7.0.9.0 and the tree was constructed with FigTree v1.3.1.

Lastly, exon-skipping was detected in all tissues. Unlike the *Drosophila* studies (see Chapters 3 and 4), the forward primers used in this study did not overlap the first base of *D. magna* exon 4 due to differences in sequences between exon variants. Therefore, a rate of exon 4-skipping could be estimated. In gut tissue, exon 4-skipping was detected at a frequency of 1.5×10^{-2}, and exon 6-skipping was double that of exon 4-skipping in guts at a frequency of 3×10^{-2}. Exon 6-skipping in guts appeared to be relatively common in *D. magna*

and appeared to occur at a frequency higher than that of the expression of several exon 6 variants in the same tissue (Fig. 2). In carcasses exon 4-skipping was detected at a frequency of $3.9x10^{-2}$ and was comparable to the expression of exon 4.7 in the same tissue, and exon 6-skipping at a frequency of $3x10^{-2}$, again occurring at a frequency higher than several exon 6 variants (Fig. 2). Finally, exon 4-skipping detected in haemolymph was at a frequency of $5.6x10^{-4}$, while exon 6-skipping in haemolymph was detected at a frequency comparable to the frequencies seen in the other tissue types ($3.3x10^{-2}$). Thus, exon 4- and exon 6-skipping appeared to be relatively common in *D. magna Dscam*. Exon 4-skipping was proportionally higher in carcasses than in guts or haemolymph, and exon 6-skipping was similar in all tissues.

5.3.4. Detecting differences in constitutive exon variant expression between tissues

I used two methods of analysis to examine constitutive expression between tissues in control *D. magna*. I used DESeq to test the number of reads for each *Dscam* exon 4 and 6 variant for being differential between tissue types. I found no difference in the relative expression of any exon 4 or 6 variant between guts and carcasses (exon 4: all q-values = 1; exon 6: all q-values $\geq$ 0.99), guts and haemolymph (exon 4: all q-values $\geq$ 0.58; exon 6: all q-values $\geq$ 0.91), or between carcasses and haemolymph (exon 4: all q-values = 1; exon 6: all q-values $\geq$ 0.92). Analysis with MCMCglmm however revealed a significant exon 4-by-tissue effect in models including and excluding haemolymph treatments (Tables 1 and 2). Specifically, some 2.5% of the total variance explained in the models was attributable to an exon 4-by-tissue effect (Tables 1 and 2). Furthermore, small exon 6-by-tissue effects were also seen in both models (Tables 1 and 2). Thus, subtle tissue-specific exon 4- and 6-expression may occur in *Daphnia* but the different methods of analyses were not consistent.

5.3.5. D. magna Dscam expression in response to different parasite genotypes

I estimated indices of diversity, and used DESeq and MCMCglmm to test whether Dscam expression in *D. magna* was altered in response to different parasite-exposures. Firstly, I estimated tissue-specific Dscam expression diversity for each treatment in the study (see Table S7). Figure 5 shows the mean Simpson's Index of Diversity values estimated for exon 4, exon 6, and exon 4-6 combinations for each tissue and each immune challenge.

There was a significant effect of tissue on exon 4 diversity ($F_{2,68}$ = 19.34, p = <0.001), but no effect of parasite ($F_{2,68}$ = 0.05, p = 0.949) or tissue-parasite interaction ($F_{2,67}$ = 0.07, p = 0.991) was evident on exon 4 diversity. For exon 6 diversity, a significant effect of tissue was also found (tissue: $F_{2,68}$ = 44.96, p = <0.001), but no effect of parasite ($F_{2,68}$ = 0.52, p = 0.597) or interaction between tissue and parasite ($F_{2,67}$ = 0.52, p = 0.721) was found. Finally, a significant tissue-effect was found for exon 4-6 combination diversity ($F_{2,68}$ = 38.51, p = <0.001), but there were no effects of parasite ($F_{2,68}$ = 0.22, p = 0.800) or tissue-parasite interaction ($F_{2,67}$ = 0.22, p = 0.924) (Fig. 5). Thus, while between-tissue differences in expression diversity were evident, no significant effects of parasite-exposure on expression diversity were detected.

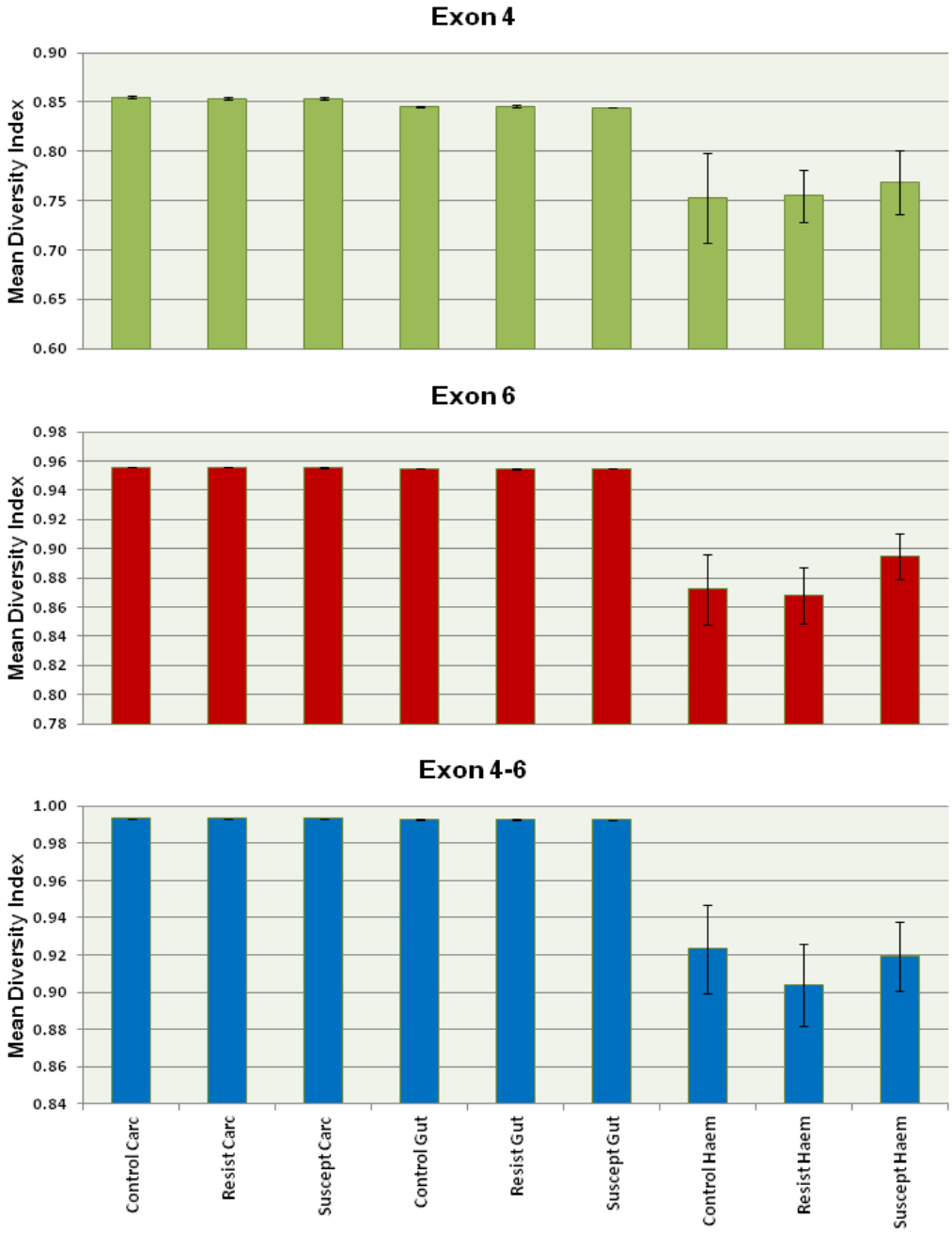


Figure 5: Expression diversity for exons 4, 6, and 4-6 combinations between tissues and between treatments. I fitted a GLM to diversity data as: (1-D) ~ tissue + parasite + tissue*parasite. Significant tissue effects were found, but no effects of parasite were evident for exon 4, exon 6, or exon 4-6 combination expression diversity. Diversity is represented as mean 1-D. Error bars represent SEM calculated from eight independent experiments. **Key**: Carc: carcass, Haem: haemolymph, Resist: resistant, Suscept: susceptible, Combo: exon 4-6 combinations.

Using MCMCglmm, I fitted a model to the count data to test for an interaction between Dscam expression and experimental conditions. Table 1 outlines the output representing the variance structure for the model. The model mixed well and the average effective sample size was around 7500 (see appendix, figure S14).

Table 1: Variance structure summarising MCMCglmm analysis of a model testing for effects of pathogen-exposure on Dscam expression. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Replicate	0.453	0.237	0.070	[0, 1.59]	1.0
Primer-Pair	**3.322**	**3.248**	**3.138**	**[2.13, 4.61]**	**7.3**
Ex4	**18.168**	**14.804**	**11.939**	**[3.33, 40.66]**	**40.0**
Ex6	**3.232**	**3.014**	**2.595**	**[1.38, 5.52]**	**7.1**
Ex4:Ex6	**0.922**	**0.915**	**0.900**	**[0.69, 1.17]**	**2.0**
Ex4:Tissue	**1.044**	**0.924**	**0.712**	**[0.27, 2.13]**	**2.3**
Ex4:Parasite	0.054	0.029	0.001	[0, 0.20]	0.1
Ex6:Tissue	**0.313**	**0.302**	**0.284**	**[0.17, 0.50]**	**0.7**
Ex6: Parasite	0.013	0.007	0.000	[0, 0.043]	0.0
Ex4:Tissue: Parasite	**0.512**	**0.492**	**0.440**	**[0.26, 0.78]**	**1.1**
Ex6:Tissue: Parasite	0.025	0.017	0.000	[0, 0.08]	0.1
Ex4:Ex6: Parasite	0.007	0.003	0.000	[0, 0.03]	0.0
Ex4:Ex6: Tissue	0.027	0.016	0.000	[0, 0.09]	0.1
Ex4:Ex6: Tissue: Parasite	0.007	0.003	0.000	[0, 0.03]	0.0
Residuals	17.297	17.297	17.314	[16.91, 17.67]	38.1

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

Much of the variance was explained by non-random exon 4, exon 6, and exon 4-6 combination expression. In addition, 2.3% of the total variance explained by random effects in the model was attributable to an exon 4-by-tissue effect, with a smaller exon 6-by-tissue effect also detected (0.7%). An exon 4-by-tissue-by-parasite effect was also seen, representing 1.1% of the total variance. The credibility bounds of each of these effects did not reach zero. Surprisingly, a relatively large amount of variance was attributable to primer-

pair, representing some 7.3% of the total variance. Significant effects caused by primer-pairs in this analysis were associated with several haemolymph treatments, but not any gut or carcass treatments (data not shown). Accordingly, after examining these results, and also discovering that expression patterns obtained from haemolymph were highly variable between biological replicates in heat maps representative of expression (see figure 6 below), the analysis was repeated with haemolymph treatments excluded. Table 2 outlines the output representing the variance structure for the model excluding haemolymph treatments. The model mixed well and the average effective sample size was around 7000 (e.g. see appendix, figure S15).

Table 2: Variance structure summarising MCMCglmm analysis of a model testing for effects of pathogen-exposure on Dscam expression. Haemolymph treatments were excluded from this analysis. The variance components are of random effects. **Key**: RE = relative expression; cub_PC = cube-root of percentage; C.I. [L/U] = Lower/Upper highest posterior density credibility intervals.

Dscam RE (cub_PC)	Posterior Mean	Posterior Median	Posterior Mode	C.I. [L/U][1]	% Total Variance[2]
Replicate	0.001	0.001	0.000	[0, 0.004]	0
Primer-pair	0.004	0.004	0.003	[0, 0.007]	0
Ex4	**19.977**	**16.235**	**11.726**	**[4.67, 45.52]**	**76.0**
Ex6	**3.918**	**3.693**	**3.311**	**[1.84, 6.50]**	**14.9**
Ex4:Ex6	**0.904**	**0.897**	**0.894**	**[0.72, 1.12]**	**3.4**
Ex4:Tissue	**0.707**	**0.541**	**0.340**	**[0.13, 1.72]**	**2.7**
Ex4:Parasite	0.001	0.000	0.000	[0, 0.004]	0
Ex6:Tissue	**0.030**	**0.027**	**0.027**	**[0.008, 0.07]**	**0.1**
Ex6: Parasite	0.000	0.000	0.000	[0, 0.001]	0
Ex4:Tissue: Parasite	**0.016**	**0.015**	**0.014**	**[0.008, 0.03]**	**0.1**
Ex6:Tissue: Parasite	0.000	0.000	0.000	[0, 0.001]	0
Ex4:Ex6: Parasite	0.000	0.000	0.000	[0, 0.001]	0
Ex4:Ex6: Tissue	**0.073**	**0.072**	**0.073**	**[0.05, 0.09]**	**0.3**
Ex4:Ex6: Tissue: Parasite	0.000	0.000	0.000	[0, 0.001]	0
Residuals	0.646	0.646	0.647	[0.63, 0.66]	2.5

[1] Credibility intervals less than 10^{-3} are rounded to zero.
[2] Percentage of the total variance explained by random effects in the model.

Here, non-random expression of exon 4, exon 6, and exon 4-6 combination expression was detected. The effect of primer-pair seen previously (Table 1) disappeared when haemolymph treatments were excluded from the analysis. In addition, tissue effects were again seen and included an exon 4-by-tissue effect which represented 2.7% of total variance explained by random effects in the model. Moreover, a very small exon 6-by-tissue effect (0.1%), and an exon 4-6 combination-by-tissue effect (0.3%) was apparent. The credibility bounds of each of these effects did not reach zero but the lower bounds were very close to zero. Notably, a very small exon 4-by-tissue-by-parasite effect (0.1%) was also apparent. Thus, effects of tissue were detected as before (Table 1), but only a weak effect of parasite-exposure was evident when haemolymph treatments were excluded from the analysis.

I then used DESeq and MCMCglmm to test for an effect of parasite treatment on the Dscam splicing pattern. Figure 6 shows heat maps representative of exon 4- and exon 6-expression for all treatments in the study. Strong patterns were seen for exon variants across treatments for both exons 4 and 6. For example, exon 4 variants 4.1, 4.4 and 4.8 were more abundant across treatments relative to other exon 4 variants. However, visual inspection suggested that no discernible patterns were obvious between different treatments.

I also investigated whether patterns of expression of exon 4-6 *combinations* were apparent between tissues and between treatments. Figure 7 shows a heat map of the relative expression of exon 4-6 combinations of which all represented over 0.5% in at least one treatment in the study.

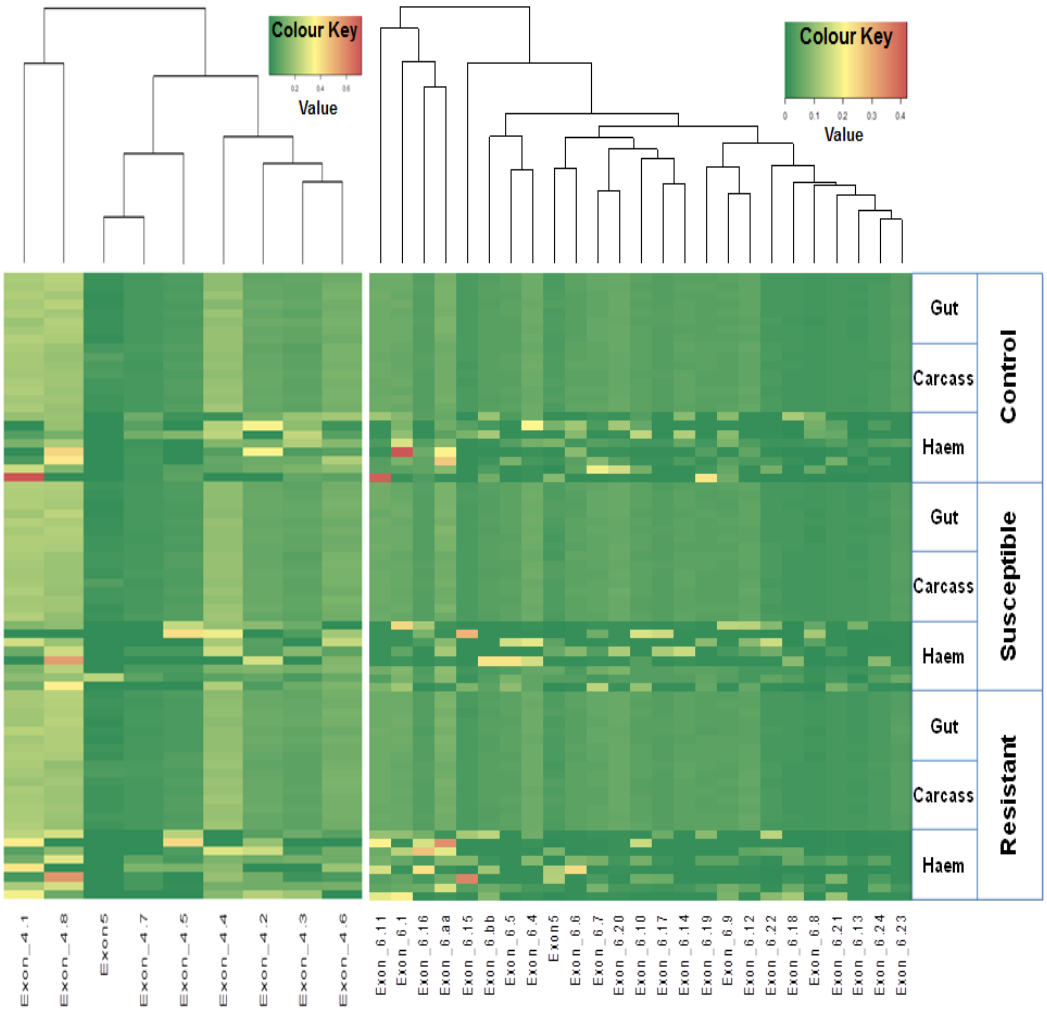


Figure 6: Heat map representation of relative expression of exon 4 (left block) and exon 6 (right block) variants for all treatments in this study. Patterns of expression for individual variants of both exons 4 and 6 across treatments were evident. For example, exons 4.1, 4.4 and 4.8 were the most abundantly expressed exon 4 variants in all treatments. Notably exon 4 and exon 6 expression in haemolymph appeared to show more variation in proportional expression of variants between biological replicates. No obvious patterns of expression were detected between treatments. The inclusion of exon 5 indicates relative abundance of exon-skipping for each exon cluster. Each tissue type, i.e. Gut, Carcass, Haem (haemolymph) included 8 biological replicates, and each cell represents the proportional expression of a given exon variant in a single sample. Dendrograms result from hierarchical clustering of exon variant expression pattern using Euclidean distance measures.

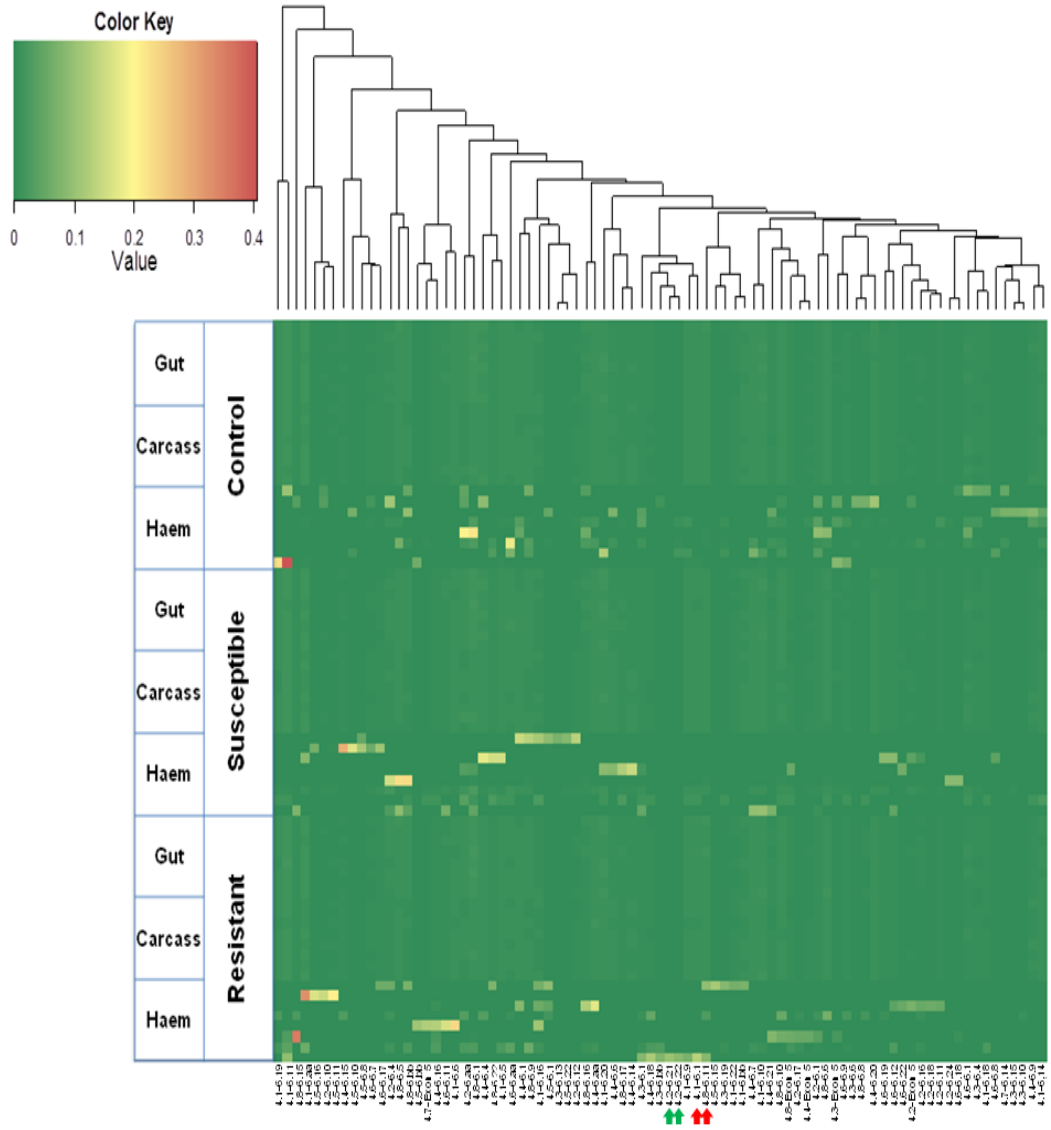


Figure 7: **Heat map representation of relative expression of exon 4-6 combinations of which each represented a minimum of 0.5% in at least one treatment in this study.** Patterns of expression for exon 4-6 combinations across treatments were evident. For example, transcripts containing the combinations 4.1-6.1 and 4.8-6.11 (red arrows) were higher in abundance across all treatments than transcripts containing the combinations 4.2-6.21 and 4.2-6.22 (green arrows). Notably, exon 4-6 combination expression in haemolymph showed more variation in proportional expression of many combinations between biological replicates. Each cell represents the proportional expression of a given exon combination in a single sample. Dendrograms result from hierarchical clustering of exon combination expression pattern using Euclidean distance measures.

As with individual exons, visual inspection suggested that patterns were evident across treatments. For example, transcripts containing the combinations 4.1-6.1 and 4.8-6.11 were higher in abundance across all treatments than transcripts containing the combinations 4.2-6.21 and 4.2-6.22. However, as with independent exon 4- and 6-expression (Fig. 6), visual inspection suggested that no discernible patterns were obvious between different treatments.

Notably, haemolymph in all treatments showed significantly more variation in exon 4- and 6-expression patterns between biological replicates (see Figures 6 and 7). For example, patterns of higher variation in exon 4-6 combination expression were seen in haemolymph compared to other tissues of *Daphnia* exposed to both strains of *P. ramosa* (Fig. 7). This variation between biological replicates was not seen in guts or carcasses. Moreover, separate examination of the 8 individual replicates for each tissue type showed high variation between them in haemolymph, but variation between biological replicates in guts and carcasses were much more similar (not shown).

Following DESeq analysis, I found no differential expression of any exon 4 or 6 variant in any tissue in response to exposure to either strain of *P. ramosa* (all q-values ≥ 0.99). Equally, no differential expression in response to pathogen-exposure was detected in any exon 4-6 combination in any tissue (all q-values = 1). Following MCMCglmm analysis inclusive of haemolymph treatments, I found that exon 4.1 in haemolymph appeared to be affected by strain *r* of *P. ramosa* ($p = 0.01$). However, in analysis excluding haemolymph treatments, no significant effects of parasite-exposure were evident.

5.4. Discussion

5.4.1. Dscam is non-randomly expressed in D. magna but tissue-regulated expression could not be confirmed

Constitutive expression in Daphnia

As seen in *Drosophila* (see Chapter 4), variants of exons 4 and 6 are non-randomly expressed in all tissues examined in *D. magna*. This was observed as strong deviation from even expression, and is indicative of some level of regulation (Fig. 2). In addition, analysis with MCMCglmm confirmed that effects of exon 4, exon 6, and exon 4-6 combinations made significant contributions to total variance found in the model whether haemolymph treatments were included or excluded (see Tables 1 and 2). The results are also partly consistent with previous work on *D. magna* Dscam. Brites et al (2008) observed that patterns of expression of exons 4, 6 and 11 strongly deviated from random expectation in haemocytes.

Patterns of combinatorial splicing of exons 4 and 6 were comparable to those detected in *Drosophila* (see Chapters 3 and 4). Exon 4-6 combinations in control *Daphnia* appeared to be close to random in guts and carcasses, with some combinations detected at frequencies deviating from random expectation (Fig. 3). MCMCglmm analysis including haemolymph treatments found no significant exon combination-by-tissue effects (Table 1). However, analysis excluding haemolymph treatments detected a significant exon combination-by-tissue effect (Table 2). Following analysis with DESeq, no significant differences in exon combination expression between tissues was detected. In contrast to gut and carcass tissues, exon-exon splicing in haemolymph appeared to be much more variable (Fig. 3). Separate examination of the 8 individual replicates showed high variation between them in haemolymph, but not in other tissues. Moreover, a high level of variance was attributed to primer-pair in the GLMM including haemolymph treatments (Table 1) and these effects

were caused only by primer-pairs associated with haemolymph treatments. Thus, effects of early-round PCR stochasticity may have occurred, but that high variation exists in Dscam expression in haemolymph at an individual level cannot be excluded.

Unlike expression in *Drosophila*, where some exon 4 variants appeared to show tissue-specific expression (see Chapter 4), expression of exon 4 and 6 variants appears to be similar between tissues in *Daphnia*. However subtle tissue-specific expression in the crustacean could not be excluded. While I could detect no difference in the relative expression levels of exon 4 and 6 variants between tissues with DESeq analysis, I found significant exon-by-tissue effects for both exons 4 and 6 using MCMCglmm (Table 1). Moreover, a model excluding the highly variable haemolymph treatments also found a significant exon-4-by-tissue effect (Table 2). In the latter analysis, the exon-6-by-tissue effect also remained, however the effect was much smaller with the lower credibility bound very close to zero.

Exon-skipping occurs at relatively high frequencies in D. magna

Exon 4- and 6-skipping was detected in *D. magna*. Exon 4-skipping was higher in carcasses than in guts, and lowest in haemolymph. Contrastingly, exon 6-skipping was comparable between all tissues assayed, a pattern also seen in my *Drosophila* study (Chapter 4). Interestingly, exon-skipping for both exons appeared to occur as frequently as the expression of some exon variants, especially true of exon 6. As speculated previously (see Chapter 4, Discussion), exon-skipping may be a failure of the splicing mechanism or possibly part of a regulatory mechanism. Based on the present data, it appears to show tissue-specific occurrence in exon 4, but not in exon 6. This could be indicative of different regulation of splicing for each exon, or a difference in splicing-fidelity between the exons. Interestingly, the sequencing motifs thought to be involved in alternative splicing of exons 4 and 6 in *Drosophila* (Graveley 2005), have also been discovered in *Daphnia* (Brites, McTaggart et al.

2008). Thus, the splicing mechanism may be preserved between the diverged species and the purpose (or consequence) of exon-skipping may be comparable between the species.

Novel D. magna Dscam alleles were discovered

All expected exon 4 variants were found in this study, and all expected exon 6 variants were also found with the exception of exons 6.2 and 6.3. However, I discovered two variants of exon 6 which did not match the reference genome, designated 6.aa and 6.bb. Following phylogenetic tree construction based on maximum likelihood method (Fig. 4), I found the amino acid sequence of variant 6.bb to be closest to the classical exon 6.3. Interestingly, previous work in *D. magna* Dscam failed to detect exon 6.3 (Brites, McTaggart et al. 2008). Thus, 6.bb is very likely an allelic variation of the reference exon 6.3. Exon 6.aa however, although most similar to exon 6.17, did not closely match the amino acid sequence of any classical *D. magna* exon 6 variant. That 6.aa is a novel exon 6 variant cannot be excluded. However it should be noted that only a part of the exon sequence was available for clustering analysis and a full-length sequence would be required for a more accurate comparison.

5.4.2. D. magna Dscam expression in response to natural parasite diversity

I investigated whether *D. magna* Dscam gene expression was affected by exposure to different isolates of *P. ramosa*. I used three methods of analysis as previously employed in Chapters 3 and 4, and found only weak effects of parasite-exposure on Dscam expression in *D. magna*. A small exon 4-by-tissue-by-parasite effect was detected using a MCMCglmm model inclusive of haemolymph treatments (Table 1), however analysis excluding the highly variable haemolymph treatments found that the effect was dramatically reduced, and although credibility bounds did not reach zero, the lower bound was very close (Table 2). Moreover, following DESeq analysis, no differential expression of variants of exon 4 or 6, or combinations thereof, were detected in response to any strain of *P. ramosa*. The calculation of an FDR in one but not both of the complementary analyses may have contributed to the

small differences seen. Additionally, a GLM fitted to estimates of expression diversity found no effects of parasite-exposure. This suggested that either no change in expression diversity occurs in *D. magna* in response to the parasites used in this study, or that the strength of the analysis is too weak to detect any changes. This reflects expression diversity analysis seen in *Drosophila* (Chapter 4), which also showed no effect of pathogen-exposure in any tissue.

Interestingly, Brites (2010) also examined Dscam exon 4-expression between two different isolates of *P. ramosa* (one of which the host was susceptible to infection, one of which the host was resistant) using qRT-PCR and also found no supporting evidence that regulation of alternative Dscam exons occurred in response to parasite-exposure. Additionally, using qRT-PCR to measure levels of overall Dscam transcripts in response to parasite-exposure, no change in total Dscam transcript abundance was found in either resistant or susceptible hosts (Brites 2010). Incidentally, this was also consistent with studies in *Anopheles*, where no overall changes of Dscam transcript abundance accompanied parasite-exposure, indicating that Dscam transcripts are not up-regulated in response to parasites (Dong, Taylor et al. 2006). Brites (2010) concedes that the critical time point to detect such effects may have been missed, having harvested *Daphnia* one week post-exposure. Nonetheless, the present study and an independent *Daphnia* study have found no strong supporting evidence of effects of parasite-exposure on the usage of alternative exons 4 or 6 in *D. magna*.

It was possible that changes in Dscam expression were missed because the wrong time point (4 hours) was chosen. Interestingly, an independent study in another crustacean, *P. leniusculus*, reported the significant induction of Dscam mRNA expression post-exposure to bacteria after 6 hours. Nevertheless, as a cellular immune response has been previously shown in *Daphnia* at 4 hours (Auld, Scholefield et al. 2010), Dscam-mediated expression to parasite-exposure could have been expected at this time.

Finally, despite the use of a cloned host-line of synchronised individuals, the haemolymph in all treatments showed significantly more variation in exon 4- and 6-expression (Fig. 6), and many exon 4-6 combinations (Fig. 7) between biological replicates. Whether the source of this variation originated from early-round PCR stochasticity, originated at the level of RNA extraction or cDNA synthesis, or is a true reflection of Dscam expression in *D. magna* haemolymph, is unknown. It will be important to increase as much as possible the amount of mRNA extracted from haemolymph to ensure samplings of the cDNA pool can be reliably representative of Dscam expression in this tissue type. Future studies may benefit from the use of flow cytometry and fluorescence-activated cell sorting to both purify and quantify immune cells prior to high-throughput sequencing.

5.5. Conclusion

It appears that although a significant cellular response can occur in *Daphnia* following exposure to natural parasites (Auld, Scholefield et al. 2010), it is not accompanied by any discernible change in Dscam expression. The cellular response which occurs in susceptible hosts, but not in resistant hosts, could indicate that resistance is borne from prevention of parasite-entry rather than triggered immune mechanisms (Auld, Scholefield et al. 2010). As such, one could have expected detectable differences in Dscam expression patterns between the exposures where one parasite drives gene expression while the other has little or no effect, however this was not apparent. In the present study and in an independent study on *Daphnia* Dscam expression (Brites 2010), no obvious effects of parasite-exposure were seen on Dscam exon 4- or 6-expression, or indeed on the total abundance of Dscam transcripts (see Brites 2010).

To date, I have discovered subtle exon 4-by-challenge effects in independent experiments with *Anopheles* (Chapter 2), and in *Drosophila* (Chapters 3 and 4), with a very small effect here in *Daphnia* (Table 2). If Dscam is involved in the *Daphnia* response to isolates of *P. ramosa* it could be constitutive, in other words, not influenced further by immune challenge. Future work comparing Dscam expression between different host genotypes in response to natural parasites may reveal differences in constitutive Dscam expression which could be influencing parasite infection success.

6. General Discussion

This chapter was written with comments from Darren Obbard and Tom Little.

6.1. Pathogen-specificity in immunity

In vertebrates, an understanding of pathogen-specificity is well established. Through antibody diversification, vertebrates can recognise and respond discriminately to millions of antigens. Vertebrates also benefit from long-term pathogen-specific memory, which results in a more rapid and efficient response to repeat exposures. Whether invertebrates have pathogen-specific immunity remains uncertain, but some level of specificity in invertebrate immunity is known. Similar to vertebrates, invertebrates show specific recognition of pathogen associated molecular patterns (PAMP) through pattern recognition receptors (PRRs) (Medzhitov and Janeway 2000). For example, peptidoglycan recognition proteins (PGRPs) respond to the presence of peptidoglycan (PG), an essential and highly conserved component of the bacterial cell wall, by inducing antimicrobial effects (Dziarski and Gupta 2005). This could be described as pathogen *class*-specific immunity.

Invertebrates were thought to lack pathogen-specific immunity or specific memory primarily because they seem to lack the equivalent molecules which confer these phenomena in vertebrates, such as antibodies. Moreover, the necessity of relatively short-lived invertebrates requiring a complex immune system could be questioned. Nevertheless, an 'enhanced immunity' to repeat infection was found in invertebrates some 40 years ago (Boman, Nilsson et al. 1972), and several examples of a higher level of specificity in the immune response of invertebrates have been discovered since (Kurtz and Franz 2003; Sadd, Kleinlogel et al. 2005; Sadd and Schmid-Hempel 2006; Sadd and Schmid-Hempel 2007; Johnson, van Hulten et al. 2008), indicative of a more complex immune system than previously supposed. Furthermore, fitness benefits of invertebrates possessing an apparently complex immune system have been discovered (Little, O'Connor et al. 2003; Moret and Siva-Jothy 2003), and experimental data has even suggested that specific memory exists in an invertebrate (Pham, Dionne et al. 2007). One could anticipate that invertebrates such as insects, which have been

exposed to infectious micro-organisms for over 400 million years (Grimaldi 2009), could have evolved pathogen-specificity and specific memory to defend against disease. Yet, examples of conclusive pathogen-specificity or specific memory in invertebrates are few and often unconvincing.

6.2. Dscam and pathogen-specificity

Dscam is currently the best candidate for the underlying mechanism for a pathogen-specific response in invertebrate immunity. Implicated in having a role in phagocytosis, a great deal of excitement has surrounded Dscam's possible involvement in specificity in invertebrate immunity (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). However, evidence of pathogen-specific Dscam splicing is limited. So far, a pathogen-specific expression of Dscam isoforms has been reported in the mosquito and crayfish only (Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011).

In the mosquito study, quantitative RT-PCR (qRT-PCR) was used to determine the fold change in relative expression of Dscam exon 4 variants in response to different immune elicitors (Dong, Taylor et al. 2006). The study reported levels of over- and under-representation of exon variants, detecting up to 4-fold changes in either direction. The implications of infection-responsive splicing of Dscam in a malaria vector would be significant. However, qRT-PCR may not be the most reliable method to study relative Dscam exon variant-expression. For instance, different methods of analysis of qRT-PCR data can differ in their accuracy and reproducibility, and can even lead to opposing biological conclusions (Skern, Frost et al. 2005; Cikos, Bukovska et al. 2007). It has been proposed that using two or more analytical approaches would help to validate conclusions

made from qRT-PCR data (Skern, Frost et al. 2005). To my knowledge, Dong et al (2006) used only one method of analysis. Moreover, only exon 4 variant-expression was assayed in the mosquito, and only three replica assays were carried out to determine fold-changes of expression in the qRT-PCR analysis. Qualitative information gathered from the splicing of two or more alternative exons (e.g. variable exons 4, 6 and 10 in *A. gambiae*), and an increase of biological replication may have led to more robust or even different results.

Both *Anopheles* and *Pacifastacus* studies do show evidence of pathogen-specific binding of some exon variants (Dong, Taylor et al. 2006; Watthanasurorot, Jiravanichpaisal et al. 2011). In the *Anopheles* study, bacterial binding assays showed that the affinity of immune challenge-responsive Dscam exon 4 variants correlated with their induction specificity determined by qRT-PCR (Dong, Taylor et al. 2006). For example, exon 4.8 appeared to be significantly over-expressed by cells challenged with *E. coli* and splice forms containing exon 4.8 had a higher binding affinity to *E. coli*. However, after a similar binding assay but with selective silencing of exon 4.8 through RNAi-targeting, decreased binding of Dscam to *E. coli* and another gram-negative bacterium, *Pseudomonas veronii*, was seen. In addition, expression patterns generated by these two gram-negative bacteria were very similar in challenged Sua5B cells (Pearson's correlation coefficient, $r = 0.80$), while cells exposed to lipopolysaccharide (characteristic of gram-negative bacteria) or peptidoglycan (characteristic of gram-positive bacteria) showed only some degree of correlation (data not given). This is suggestive of pathogen *class*-specific splice-form expression. In the crayfish study, different representatives of gram-negative or gram-positive bacteria were not compared (Watthanasurorot, Jiravanichpaisal et al. 2011). Therefore, it could not be determined whether pathogen-specific rather than pathogen *class*-specific Dscam expression patterns and binding activity occurs in this crustacean. While both studies showed that Dscam mediates in phagocytosis, and is associated with immune function, they did not show definitive pathogen-specific Dscam splicing.

In this thesis, I examined the Dscam splicing pattern in response to immune challenge using a powerful sequencing method. The hypervariable form of the *Dscam* gene has never been found outside the Pancrustacea (Crayton, Powell et al. 2006; Armitage, Freiburg et al. 2012). Accordingly, I examined infection-responsive Dscam expression in three host species representative of both the insects and crustaceans.

6.3. Findings of the present study

In Chapter 2, I followed on from the work of Dong et al (2006) where pathogen-specific splicing of Dscam was first reported in the *Anopheles* mosquito. Using a cloning and sequencing approach, and examining Dscam expression diversity in the mosquito in response to different genotypes of its natural parasite, *P. falciparum*, I demonstrated that diversity increased in parasite-exposed mosquitoes. However, I found only weak evidence that diversity increased further in response to increasing parasite diversity. I also discovered that although it was the combined diversity of exons 4 and 6 that drove the patterns seen in both the field and the laboratory, exon 4 had a close-to-significant effect on the relationship between expression diversity and parasite diversity.

In Chapter 3, I extended the study to a more experimentally amenable organism and adopted a much more powerful method of approach. Using a specially-designed Illumina-based sequencing method, I examined both constitutive and infection-responsive Dscam expression in *Drosophila*. I confirmed that Dscam exons 4 and 6 are non-randomly expressed, indicative of some level of regulation, and that the method was capable of detecting allelic variations of exon variants and novel exons of *Drosophila* Dscam. Interestingly, I also found a small but significant exon 4-by-challenge effect. However, the pathogen-induced effects on Dscam

expression in *Drosophila* analysed with deep sequencing technology fell short of the striking results previously reported in *Anopheles* (Dong, Taylor et al. 2006).

As the background of the whole fly may have obscured pathogen-induced effects on the expression of *Dsca*m, I next examined expression in separated tissues in response to immune elicitors of *Drosophila* (Chapter 4). I reasoned that analysing tissue-specific patterns of expression should reduce the nervous-system background. As such, the signal of an immune-related pattern would be less obscured, especially in response to well-characterised immune elicitors such as the bacteria *Escherichia coli* and *Micrococcus luteus*. I again found small pathogen-induced effects on Dscam exon 4-expression, which also appeared to be dependent upon tissue. Moreover, results of the high-throughput sequencing approach confirmed that exon 4 was expressed in a tissue-specific manner, and determined that unlike exon 4-expression, exon 6-expression did not appear to be tissue-dependent.

Finally, in Chapter 5 I switched to another member of the Pancrustacea, namely the crustacean *Daphnia*. By exposing a clonally propagated host to different strains of its natural parasite to which different infection outcomes had been previously reported (Carius, Little et al. 2001), I hoped to determine whether significant differences in Dscam expression could be seen between *Daphnia* exposed to the different challenges. Again, I found a small tissue-dependent effect on the expression of Dscam exon 4 in response to parasite-exposure. I also determined that exons 4 and 6 in *Daphnia* Dscam were non-randomly expressed in the different tissues assayed, and found that tissue-specific expression appears to be more subtle compared to that seen in *Drosophila*.

The results outlined in this thesis cannot rule out pathogen-specific Dscam splicing in the three different species used in this study, but they consistently suggest that effects, if they exist at all, are extremely small. As such, the possibility that their significance is statistical

artefact cannot be ruled out either. Interestingly however, all detected effects appeared to be attributable only to the exon 4 cluster.

Comparison with other work

What may have led to the differences seen between the present work and others? Dong et al (2006) reported pathogen-specific Dscam splicing in *Anopheles* after exposure to an array of immune elicitors in a mosquito cell line. It is possible the use of cell culture provides more power as it eliminates the background of the whole insect. However, the same study also reported infection-responsive splicing in adult mosquitoes in response to different types of *Plasmodium* (Dong, Taylor et al. 2006). My work included an estimation of the diversity of expression of many possible Dscam isoforms, and in whole insects exposed to different treatments (Smith, Mwangi et al. 2011). This turned out to be a relatively weak form of analysis, seemingly lacking the capacity to detect the pathogen-specificity previously reported of *Anopheles* Dscam (Dong, Taylor et al. 2006). Should the high-throughput sequencing method developed for the *Drosophila* and *Daphnia* studies (Chapters 3-5) have been employed to study Dscam in *Anopheles* (Chapter 2), results may have been different. Nevertheless, a failure to find definitive infection-responsive Dscam splicing in *Drosophila* and *Daphnia* is interesting. It is possible that *Anopheles* mosquitoes benefit from having Dscam-mediated pathogen-specificity but other invertebrates do not. After all, profound differences in the lifestyles and environmental exposures exist between *Anopheles*, *Drosophila*, and *Daphnia*. However, pathogen-specific splicing has also been reported in the crayfish, and therefore may be apparent in both insect and crustacean species. As such, pathogen-specific Dscam expression in *Drosophila* and *Daphnia* could also have been expected.

6.4. Future directions

Is Dscam hyper-diversity important in immunity?

How important the hyper-diversity of Dscam is within the context of invertebrate immunity remains uncertain. Does it confer a greater capacity for pathogen-recognition? Is the nature of the diversity what matters, or is arbitrary diversity enough for Dscam's role in the immune system? It has been established that although individual Dscam-expressing neurons are required to express distinct isoforms, the specific identity of the isoforms expressed in an individual nerve cell does not appear to be important (Hattori, Demir et al. 2007). Brites (2010) has proposed that Dscam may simply confer cell-to-cell recognition of haemocytes in order to prevent aggregation. In other words, the gene may have comparable roles in both the nervous and immune systems.

Measuring cell function after gene silencing

It has been shown that by silencing Dscam, significant reduction in phagocytosis is seen in different species (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006). This strongly suggests that Dscam may have a role in the pathogen-recognition and destruction process. However, if Dscam is otherwise involved in general cellular functions, such as cellular extensions, its silencing may simply render a cell less competent at many biological processes, including phagocytic uptake of pathogens. Additionally, it is entirely feasible that Dscam plays a role in immunity but does so in cooperation with other molecules serving common functions such as pathogen-clearance. For example, Dscam may be part of a multi-protein complex on the surface of haemocytes, but carries out a cellular function important for phagocytosis but not directly related to pathogen-specificity. As such, relating loss-of-function directly to a particular silenced gene may be useful in elucidating the mechanism of a cellular process such as pathogen-recognition, but it cannot determine exactly what part the gene plays in that process without more in-depth examination.

Further questions

A better understanding of the constitutive expression dynamics of Dscam may reward subsequent approaches to investigate infection-responsive expression. For example, it is currently uncertain how the expression of Dscam mRNAs and protein isoforms change over time. As such, it would be extremely useful to determine the approximate half-life of Dscam mRNAs (Neves, Zucker et al. 2004). Methods such as pulse-chase analysis (Jansens and Braakman 2003) may allow the incorporation of a labelled compound into Dscam molecules to help determine not only the movement of Dscams within a cell, but exactly for how long Dscams are expressed and under different conditions. This method would be of particular interest to monitor Dscam isoforms in an adherent cell-line which both expresses Dscam and undergoes phagocytosis.

Soluble forms of Dscam have been found in *Anopheles* and *Drosophila* haemolymph (Watson, Puttmann-Holgado et al. 2005; Dong, Taylor et al. 2006), and may be produced by proteolytic cleavage of the membrane-bound forms (Watson, Puttmann-Holgado et al. 2005). Secreted isoforms circulating in haemolymph are unlikely to be involved in neural activity (Brites 2010), and they have been proposed to play a role in opsonisation, a process which facilitates phagocytosis by recruiting antigens to the surfaces of haemocytes (Watson, Puttmann-Holgado et al. 2005). An assessment of the expression and function of soluble forms of Dscam may help elucidate how Dscam mediates phagocytosis. For example, determining whether soluble Dscams bind to the surface of phagocytic cells may strengthen the hypothesis that they act as opsonins. Moreover, measuring phagocytic activity in the presence and absence of soluble Dscams may give an indication of their relative importance in pathogen-clearance.

Lastly, the most striking report of Dscam splicing within the context of immunity originated in *Anopheles* (Dong, Taylor et al. 2006). Thus, a natural next step may be to investigate Dscam splicing in the haemocyte-like immune competent Sua5B mosquito cell-line in response to different species of *Plasmodium* but applying high-throughput sequencing technology. This would enable the quantification of the expression pattern, exclude the background of the whole insect or insect tissues, and allow comparison between the independent studies.

6.5. Concluding remarks

This thesis introduces a novel Illumina-based sequencing method useful for the quantification of the Dscam expression pattern and analysis of complex alternative splicing of *Dscam* in different organisms under different conditions. Prospects for its adoption include the investigation of Dscam splicing in more insect and crustacean species where the hypervariable form of the gene has been found (e.g. *Anopheles*, *Apis*, *Pacifastacus*, *Litopenaeus*), and the examination of Dscam expression in both the nervous and immune systems. Finally, the implications of pathogen-specific recognition in invertebrates could be huge, but I found only weak evidence that Dscam responds in a pathogen-specific manner in *Anopheles*, *Drosophila* and *Daphnia*. When examining expression of exon 4 in the three different species, I could not rule out effects of immune challenge on the expression of variants of this exon cluster. Specifically, small exon 4-by-challenge effects were seen in independent studies of all three tested host species. Thus, the possibility remains that a Dscam-mediated infection-responsive splicing mechanism may one day be revealed.

Citations

Afrane, Y. A., T. J. Little, et al. (2008). "Deforestation and vectorial capacity of *Anopheles gambiae* giles mosquitoes in malaria transmission, Kenya." Emerging Infectious Diseases **14**(10): 1533-1538.

Akira, S. and K. Takeda (2004). "Toll-like receptor signalling." Nature Reviews Immunology **4**(7): 499-511.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biology **11**(10): 1-12.

Anderson, T. J. C., B. Haubold, et al. (2000). "Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*." Molecular Biology and Evolution **17**(10): 1467-1482.

Anderson, T. J. C., X. Z. Su, et al. (1999). "Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples." Parasitology **119**: 113-125.

Armitage, S. A., R. Y. Freiburg, et al. (2012). "The evolution of Dscam genes across the arthropods." Bmc Evolutionary Biology **12**: 53.

Auld, S., J. A. Scholefield, et al. (2010). "Genetic variation in the cellular response of Daphnia magna (Crustacea: Cladocera) to its bacterial parasite." Proceedings of the Royal Society B-Biological Sciences **277**(1698): 3291-3297.

Baintner, K. (2007). "Transmission of antibodies from mother to young: Evolutionary strategies in a proteolytic environment." Veterinary Immunology and Immunopathology **117**(3-4): 153-161.

Barnes, A. I. and M. T. Siva-Jothy (2000). "Density-dependent prophylaxis in the mealworm beetle Tenebrio molitor L-(Coleoptera : Tenebrionidae): cuticular melanization is an indicator of investment in immunity." Proceedings of the Royal Society of London Series B-Biological Sciences **267**(1439): 177-182.

Bell, A. S. and L. C. Ranford-Cartwright (2004). "A real-time PCR assay for quantifying Plasmodium falciparum infections in the mosquito vector." International Journal for Parasitology **34**(7): 795-802.

Benjamini, Y. and Y. Hochberg (1995). "CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING." Journal of the Royal Statistical Society Series B-Methodological **57**(1): 289-300.

Beutler, B. (2004). "Innate immunity: an overview." Molecular Immunology **40**(12): 845-859.

Bhasin, V. K. and W. Trager (1984). "Gametocyte-Forming and Non-Gametocyte-Forming Clones of *Plasmodium falciparum*." American Journal of Tropical Medicine and Hygiene **33**(4): 534-537.

Blandin, S., S. H. Shiao, et al. (2004). "Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*." Cell **116**(5): 661-670.

Boehm, T. (2007). "Two in one: dual function of an invertebrate antigen receptor." Nature Immunology **8**(10): 1031-1033.

Boman, H. G., I. Nilsson, et al. (1972). "Inducible Antibacterial Defence System in *Drosophila*." Nature **237**(5352): 232-&.

Boots, M. and R. G. Bowers (2004). "The evolution of resistance through costly acquired immunity." Proceedings of the Royal Society of London Series B-Biological Sciences **271**(1540): 715-723.

Boulinier, T. and V. Staszewski (2008). "Maternal transfer of antibodies: raising immuno-ecology issues." Trends in Ecology & Evolution **23**(5): 282-288.

Brennan, C. A. and K. V. Anderson (2004). "Drosophila: The genetics of innate immune recognition and response." Annual Review of Immunology **22**: 457-483.

Breslow, N. E. and D. G. Clayton (1993). "APPROXIMATE INFERENCE IN GENERALIZED LINEAR MIXED MODELS." Journal of the American Statistical Association **88**(421): 9-25.

Brett, D., H. Pospisil, et al. (2002). "Alternative splicing and genome complexity." Nature Genetics **30**(1): 29-30.

Brites, D. (2010). Evolution and Expression of the Highly Variable Cell Adhesion Molecule Dscam in the Crustacean *Daphnia* and other Arthropods, University of Basel. **PhD Thesis**.

Brites, D., F. Encinas-Viso, et al. (2011). "Population genetics of duplicated alternatively spliced exons of the dscam gene in Daphnia and Drosophila." PLoS One **6**(12): e27947.

Brites, D., S. McTaggart, et al. (2008). "The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like in insects." Molecular Biology and Evolution **25**(7): 1429-1439.

Bulet, P., S. Cociancich, et al. (1992). "A Novel Insect Defensin Mediates the Inducible Antibacterial Activity in Larvae of the Dragonfly Aeschna-Cyanea (Paleoptera, Odonanta)." European Journal of Biochemistry **209**(3): 977-984.

Carius, H. J., T. J. Little, et al. (2001). "Genetic variation in a host-parasite association: Potential for coevolution and frequency-dependent selection." Evolution **55**(6): 1136-1145.

Carter, R., L. Ranford-Cartwright, et al. (1993). "The culture and preparation of gametocytes of *Plasmodium falciparum* for immunochemical, molecular, and mosquito infectivity studies." Methods in Molecular Biology; Protocols in molecular parasitology: 67-88.

Carvalho, G. R. and R. N. Hughes (1983). "THE EFFECT OF FOOD AVAILABILITY, FEMALE CULTURE-DENSITY AND PHOTOPERIOD ON EPHIPPIA PRODUCTION IN DAPHNIA-MAGNA STRAUS (CRUSTACEA, CLADOCERA)." Freshwater Biology **13**(1): 37-46.

Celotto, A. M. and B. R. Graveley (2001). "Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated." Genetics **159**(2): 599-608.

Chang, T. L. and M. E. Klotman (2004). "Defensins: Natural anti-HIV peptides." Aids Reviews **6**(3): 161-168.

Chen, B. E., M. Kondo, et al. (2006). "The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila." Cell **125**(3): 607-620.

Cherry, S. and N. Silverman (2006). "Host-pathogen interactions in drosophila: new tricks from an old friend." Nature Immunology **7**(9): 911-917.

Chou, P. H., H. S. Chang, et al. (2009). "The putative invertebrate adaptive immune protein *Litopenaeus vannamei* Dscam (LvDscam) is the first reported Dscam to lack a transmembrane domain and cytoplasmic tail." Developmental and Comparative Immunology **33**(12): 1258-1267.

Christensen, B. M., J. Y. Li, et al. (2005). "Melanization immune responses in mosquito vectors." Trends in Parasitology **21**(4): 192-199.

Christophides, G. K., D. Vlachou, et al. (2004). "Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*." Immunological Reviews **198**(1): 127-148.

Cikos, S., A. Bukovska, et al. (2007). "Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis." Bmc Molecular Biology **8**.

Combes, C. (2000). "Selective pressures in host-parasite systems." Journal de la Societe de Biologie **194**(1): 19-23.

Cong, M., L. S. Song, et al. (2008). "The enhanced immune protection of Zhikong scallop *Chlamys farreri* on the secondary encounter with *Listonella anguillarum*."

Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **151**(2): 191-196.

Cooper, E. L., B. Rinkevich, et al. (1992). "Invertebrate Immunity - Another Viewpoint." Scandinavian Journal of Immunology **35**(3): 247-266.

Cooper, E. L. and P. Roch (1986). "2nd-Set Allograft Responses in the Earthworm *Lubricus terrestris* - Kinetics and Characteristics." Transplantation **41**(4): 514-520.

Cooper, M. D. and M. N. Alder (2006). "The evolution of adaptive immune systems." Cell **124**(4): 815-822.

Cote, J. F. and K. Vuori (2002). "Identification of an evolutionarily conserved superfamily of DOCK180-related proteins with guanine nucleotide exchange activity." Journal of Cell Science **115**(24): 4901-4913.

Crayton, M. E., B. C. Powell, et al. (2006). "Tracking the evolution of alternatively spliced exons within the Dscam family." Bmc Evolutionary Biology **6**.

Dimopoulos, G., D. Seeley, et al. (1998). "Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle." Embo Journal **17**(21): 6115-6123.

Dong, Y. and G. Dimopoulos (2006). "Pattern recognition diversity in the *Anopheles gambiae* innate immune system." American Journal of Tropical Medicine and Hygiene **75**(5): 291-292.

Dong, Y. M., H. E. Taylor, et al. (2006). "AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system." Plos Biology **4**: 1137-1146.

Dostert, C., E. Jouanguy, et al. (2005). "The Jak-STAT signalling pathway is required but not sufficient for the antiviral response of drosophila." Nature Immunology **6**(9): 946-953.

Du Pasquier, L. (1992). "Origin and Evolution of the Vertebrate Immune-system." Apmis **100**(5): 383-392.

Du Pasquier, L. (2001). "The immune system of invertebrates and vertebrates." Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **129**(1): 1-15.

Du Pasquier, L. (2005). "Insects diversify one molecule to serve two systems." Science **309**(5742): 1826-1827.

Du Pasquier, L. (2006). "Germline and somatic diversification of immune recognition elements in Metazoa." Immunology letters **104**(1-2): 2-17.

Duneau, D., P. Luijckx, et al. (2011). "Resolving the infection process reveals striking differences in the contribution of environment, genetics and phylogeny to host-parasite interactions." BMC Biology **9**(1): 11.

Dziarski, R. and D. Gupta (2005). "Peptidoglycan recognition in innate immunity." Journal of Endotoxin Research **11**(5): 304-310.

Ebert, D. (2005). Ecology, Epidemiology, and Evolution of Parasitism in Daphnia [Internet], Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK2036/.

Ebert, D. (2008). "Host-parasite coevolution: Insights from the Daphnia-parasite model system." Current Opinion in Microbiology **11**(3): 290-301.

Ebert, D., R. J. H. Payne, et al. (1997). The epidemiology of parasitic diseases in Daphnia. Ecological Studies; Vertical food web interactions: Evolutionary patterns and driving forces. K. B. G. V. W. Dettner. **130:** 91-111.

Ebert, D., P. Rainey, et al. (1996). "Development, life cycle, ultrastructure and phylogenetic position of Pasteuria ramosa Metchnikoff 1888: Rediscovery of an obligate

endoparasite of Daphnia magna Straus." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **351**(1348): 1689-1701.

Ebert, D., C. D. Zschokke-Rohringer, et al. (1998). "Within- and between-population variation for resistance of Daphnia magna to the bacterial endoparasite Pasteuria ramosa." Proceedings of the Royal Society of London Series B-Biological Sciences **265**(1410): 2127-2134.

Faulhaber, L. M. and R. D. Karp (1992). "A Diphasic Immune Response Against Bacteria in the American Cockroach." Immunology **75**(2): 378-381.

Fritz, J. H., S. E. Girardin, et al. (2006). "Innate immune defense through RNA interference." Science's STKE : signal transduction knowledge environment **2006**(339): pe27.

Githeko, A. and W. Ndegwa (2001). "Predicting Malaria Epidemics in the Kenyan Highlands Using Climate Data: A Tool for Decision Makers." Global Change & Human Health **2**(1): 54-63.

Graveley, B. R. (2005). "Mutuallv exclusive splicing of the insect Dscam Pre-mRNA directed by competing intronic RNA secondary structures." Cell **123**(1): 65-73.

Graveley, B. R., A. Kaur, et al. (2004). "The organization and evolution of the Dipteran and Hymenopteran Down syndrome cell adhesion molecule (Dscam) genes." Rna-a Publication of the Rna Society **10**(10): 1499-1506.

Green, J. (1974). "PARASITES AND EPIBIONTS OF CLADOCERA." Transactions of the Zoological Society of London **32**(6): 417-516.

Grimaldi, D. A. (2009). "400 million years on six legs: On the origin and early evolution of Hexapoda." Arthropod Structure & Development **39**(2-3): 191-203.

Grindstaff, J. L., E. D. Brodie, et al. (2003). "Immune function across generations: integrating mechanism and evolutionary process in maternal antibody transmission." Proceedings of the Royal Society of London Series B-Biological Sciences **270**(1531): 2309-2319.

Guisan, A., T. C. Edwards, et al. (2002). "Generalized linear and generalized additive models in studies of species distributions: setting the scene." Ecological Modelling **157**(2-3): 89-100.

Haag, C. R., S. J. McTaggart, et al. (2009). "Nucleotide Polymorphism and Within-Gene Recombination in Daphnia magna and D. pulex, Two Cyclical Parthenogens." Genetics **182**(1): 313-323.

Hadfield, J. D. (2009). "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package." Journal of Statistical Software **33**(2): 1-22.

Hall, N. (2007). "Advanced sequencing technologies and their wider impact in microbiology." Journal of Experimental Biology **210**(9): 1518-1525.

Harcombe, W. and A. A. Hoffmann (2004). "Wolbachia effects in Drosophila melanogaster: in search of fitness benefits." Journal of Invertebrate Pathology **87**(1): 45-50.

Harding, K. C., B. J. L. Hansen, et al. (2005). "Acquired immunity and stochasticity in epidemic intervals impede the evolution of host disease resistance." American Naturalist **166**(6): 722-730.

Hartman, R. S. and R. D. Karp (1989). "Short-Term Immunological Memory in the Allograft Response of the American Cockroach, *Periplaneta americana*." Transplantation **47**(5): 920-922.

Hastings, W. K. (1970). "MONTE-CARLO SAMPLING METHODS USING MARKOV CHAINS AND THEIR APPLICATIONS." Biometrika **57**(1): 97-&.

Hattori, D., E. Demir, et al. (2007). "Dscam diversity is essential for neuronal wiring and self-recognition." Nature **449**(7159): 223-U6.

Hattori, D., S. S. Millard, et al. (2008). "Dscam-Mediated Cell Recognition Regulates Neural Circuit Formation." Annual Review of Cell and Developmental Biology **24**: 597-620.

Hauton, C. and V. J. Smith (2007). "Adaptive immunity in invertebrates: a straw house without a mechanistic foundation." Bioessays **29**(11): 1138-1146.

Hedges, L. M., J. C. Brownlie, et al. (2008). "Wolbachia and Virus Protection in Insects." Science **322**(5902): 702-702.

Hedges, L. M. and K. N. Johnson (2008). "Induction of host defence responses by Drosophila C virus." Journal of General Virology **89**: 1497-1501.

Hildemann, W. H., R. L. Raison, et al. (1977). "Immunological Specificity and Memory in a Scleractinian Coral." Nature **270**(5634): 219-223.

Hing, H., J. Xiao, et al. (1999). "Pak functions downstream of dock to regulate photoreceptor axon guidance in Drosophila." Cell **97**(7): 853-863.

Hobaek, A. and P. Larsson (1990). "SEX DETERMINATION IN DAPHNIA-MAGNA." Ecology **71**(6): 2255-2268.

Hoffmann, J. A. (1995). "Innate Immunity of Insects." Current Opinion in Immunology **7**(1): 4-10.

Hoffmann, J. A. (2003). "The immune response of Drosophila." Nature **426**(6962): 33-38.

Hoffmann, J. A., F. C. Kafatos, et al. (1999). "Phylogenetic perspectives in innate immunity." Science **284**(5418): 1313-1318.

Hummel, T., M. L. Vasconcelos, et al. (2003). "Axonal targeting of olfactory receptor neurons in Drosophila is controlled by Dscam." Neuron **37**(2): 221-231.

Ip, Y. T. (2005). "Drosophila innate immunity goes viral." Nature Immunology **6**(9): 863-864.

James, S. L. (1995). "Role Of Nitric-Oxide in Parasitic Infections." Microbiological Reviews **59**(4): 533-&.

Janeway, C. A. and R. Medzhitov (2002). "Innate immune recognition." Annual Review of Immunology **20**: 197-216.

Jansens, A. and I. Braakman (2003). "Pulse-chase labeling techniques for the analysis of protein maturation and degradation." Methods in molecular biology (Clifton, N.J.) **232**: 133-45.

Johnson, K. N., M. C. W. van Hulten, et al. (2008). ""Vaccination" of shrimp against viral pathogens: Phenomenology and underlying mechanisms." Vaccine **26**(38): 4885-4892.

Jousset, F.-X., M. Bergoin, et al. (1977). "Characterization of the Drosophila C Virus." Journal of General Virology **34**(2): 269-283.

Kapun, M., V. Nolte, et al. (2010). "Host Range and Specificity of the Drosophila C Virus." PLoS One **5**(8).

Klein, J. (1989). "Are Vertebrates Capable of Anticipatory Immune Responses." Scandinavian Journal of Immunology **29**(5): 499-505.

Klein, J. (1997). "Homology between immune responses in vertebrates and invertebrates: Does it exist?" Scandinavian Journal of Immunology **46**(6): 558-564.

Kreahling, J. M. and B. R. Graveley (2005). "The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam Pre-mRNA." Molecular and Cellular Biology **25**(23): 10251-10260.

Kurtz, J. (2004). "Memory in the innate and adaptive immune systems." Microbes and Infection **6**(15): 1410-1417.

Kurtz, J. and K. Franz (2003). "Innate defence: evidence for memory in invertebrate immunity." Nature **425**(6953): 37-38.

Lavine, M. D., G. Chen, et al. (2005). "Immune challenge differentially affects transcript abundance of three antimicrobial peptides in hemocytes from the moth Pseudoplusia includens." Insect Biochemistry and Molecular Biology **35**(12): 1335-1346.

Lee, C., N. Kim, et al. (2010). "Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution." Rna-a Publication of the Rna Society **16**(1): 91-105.

Lee, C. and M. Roy (2004). "Analysis of alternative splicing with microarrays: successes and challenges." Genome Biology **5**(7).

Lemaitre, B. and J. Hoffmann (2007). The host defense of Drosophila melanogaster. Annual Review of Immunology. **25:** 697-743.

Lemaitre, B., J. M. Reichhart, et al. (1997). "Drosophila host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms." Proceedings of the National Academy of Sciences of the United States of America **94**(26): 14614-14619.

Li, Y. J., Y. Satta, et al. (1999). "Paleo-demography of the Drosophila melanogaster subgroup: application of the maximum likelihood method." Genes & Genetic Systems **74**(4): 117-127.

Little, T. J., D. Hultmark, et al. (2005). "Invertebrate immunity and the limits of mechanistic immunology." Nature Immunology **6**(7): 651-654.

Little, T. J., B. O'Connor, et al. (2003). "Maternal transfer of strain-specific immunity in an invertebrate." Current Biology **13**(6): 489-492.

Liu, H., P. Jiravanichpaisal, et al. (2007). "Phenoloxidase is an important component of the Defense against Aeromonas hydrophila infection in a crustacean, Pacifastacus leniusculus." Journal of Biological Chemistry **282**(46).

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method." Methods **25**(4): 402-408.

Longdon, B., D. Fabian, et al. (2012). "Male-killing Wolbachia do not protect Drosophila bifasciata against viral infection." BMC Microbiology **12**(Suppl 1): S8.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics. **9:** 387-402.

Matthews, B. J., M. E. Kim, et al. (2007). "Dendrite self-avoidance is controlled by Dscam." Cell **129**(3): 593-604.

Medzhitov, R. and C. Janeway (2000). "Innate immune recognition: mechanisms and pathways." Immunological Reviews **173**: 89-97.

Meijers, R., R. Puettmann-Holgado, et al. (2007). "Structural basis of Dscam isoform specificity." Nature **449**: 487-U12.

Meister, M. (2004). "Blood cells of Drosophila: cell lineages and role in host defence." Current Opinion in Immunology **16**(1): 10-15.

Michel, T., J. M. Reichhart, et al. (2001). "Drosophila Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein." Nature **414**(6865): 756-759.

Millard, S. S. and S. L. Zipursky (2008). "Dscam-mediated repulsion controls tiling and self-avoidance." Current Opinion in Neurobiology **18**(1): 84-89.

Minakawa, N., G. Sonye, et al. (2004). "Habitat characteristics of *Anopheles gambiae s.s.* larvae in a Kenyan highland." Medical and Veterinary Entomology **18**(3): 301-305.

Moret, Y. and M. T. Siva-Jothy (2003). "Adaptive innate immunity? Responsive-mode prophylaxis in the mealworm beetle, *Tenebrio molitor*." Proceedings of the Royal Society of London Series B-Biological Sciences **270**(1532): 2475-2480.

Murphy, K. M., P. Travers, et al. (2007). Janeway's Immunobiology. New York, Garland Science.

Neves, G., J. Zucker, et al. (2004). "Stochastic yet biased expression of multiple Dscam splice variants by individual cells." Nature Genetics **36**(3): 240-246.

Nwakanma, D., A. Kheir, et al. (2008). "High gametocyte complexity and mosquito infectivity of Plasmodium falciparum in the Gambia." International Journal for Parasitology **38**(2): 219-227.

Obbard, D. J., J. Maclennan, et al. (2012). "Estimating divergence dates and substitution rates in the Drosophila phylogeny." Molecular Biology and Evolution.

Obbard, D. J., J. J. Welch, et al. (2009). "Quantifying Adaptive Evolution in the Drosophila Immune System." Plos Genetics **5**(10).

Olson, S., M. Blanchette, et al. (2007). "A regulator of Dscam mutually exclusive splicing fidelity." Nature Structural & Molecular Biology **14**(12): 1134-1140.

Orita, M., Y. Suzuki, et al. (1989). "RAPID AND SENSITIVE DETECTION OF POINT MUTATIONS AND DNA POLYMORPHISMS USING THE POLYMERASE CHAIN-REACTION." Genomics **5**(4): 874-879.

Osborne, S. E., Y. S. Leong, et al. (2009). "Variation in Antiviral Protection Mediated by Different Wolbachia Strains in Drosophila simulans." Plos Pathogens **5**(11).

Oshlack, A., M. D. Robinson, et al. (2010). "From RNA-seq reads to differential expression results." Genome Biology **11**(12).

Peng, H. and J. Zhang (2009). "Commercial high-throughput sequencing and its applications in DNA analysis." Biologia **64**(1): 20-26.

Pham, L. N., M. S. Dionne, et al. (2007). "A specific primed immune response in *Drosophila* is dependent on phagocytes." Plos Pathogens **3**.

Powell, J. R. (1997). Progress and prospects in evolutionary biology: The Drosphila model.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing,. R Foundation for Statistical Computing. Vienna, Austria.

Rahman, M. M., H. L. S. Roberts, et al. (2004). "Induction and transmission of Bacillus thuringiensis tolerance in the flour moth Ephestia kuehniella." Proceedings of the National Academy of Sciences of the United States of America **101**(9): 2696-2699.

Rheins, L. A., R. D. Karp, et al. (1978). "HUMORAL IMMUNITY INDUCED IN AMERICAN COCKROACH (PERIPLANETA-AMERICANA)." American Zoologist **18**(3): 595-595.

Roth, O. and J. Kurtz (2009). "Phagocytosis mediates specificity in the immune defence of an invertebrate, the woodlouse Porcellio scaber (Crustacea: Isopoda)." Developmental and Comparative Immunology **33**(11): 1151-1155.

Roth, O., B. M. Sadd, et al. (2009). "Strain-specific priming of resistance in the red flour beetle, *Tribolium castaneum*." Proceedings of the Royal Society B-Biological Sciences **276**(1654): 145-151.

Rowley, A. F. and A. Powell (2007). "Invertebrate immune systems-specific, quasi-specific, or nonspecific?" Journal of Immunology **179**(11): 7209-7214.

Ruano, G., D. E. Brash, et al. (1991). "PCR: the first few cycles." Amplifications **7**: 1-4.

Rutschmann, S., A. C. Jung, et al. (2000). "The Rel protein DIF mediates the antifungal but not the antibacterial host defense in Drosophila." Immunity **12**(5): 569-580.

Sadd, B. M., Y. Kleinlogel, et al. (2005). "Trans-generational immune priming in a social insect." Biology Letters **1**(4): 386-388.

Sadd, B. M. and P. Schmid-Hempel (2006). "Insect immunity shows specificity in protection upon secondary pathogen exposure." Current Biology **16**(12): 1206-1210.

Sadd, B. M. and P. Schmid-Hempel (2007). "Facultative but persistent transgenerational immunity via the mother's eggs in bumblebees." Current Biology **17**(24): R1046-R1047.

Saejeng, A., M. T. Siva-Jothy, et al. (2010). "Low cost antiviral activity of Plodia interpunctella haemolymph in vivo demonstrated by dose dependent infection." Journal of Insect Physiology **57**(2): 246-250.

Schlenke, T. A. and D. J. Begun (2003). "Natural selection drives drosophila immune system evolution." Genetics **164**(4): 1471-1480.

Schmid-Hempel, P. and D. Ebert (2003). "On the evolutionary ecology of specific immune defence." Trends in Ecology & Evolution **18**(1): 27-32.

Schmucker, D. (2007). "Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring." Nature Reviews Neuroscience **8**(12): 915-920.

Schmucker, D. and B. Chen (2009). "Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes." Genes & Development **23**(2): 147-156.

Schmucker, D., J. C. Clemens, et al. (2000). "Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity." Cell **101**(6): 671-684.

Schneider, I. (1972). "CELL LINES DERIVED FROM LATE EMBRYONIC STAGES OF DROSOPHILA-MELANOGASTER." Journal of Embryology and Experimental Morphology **27**(APR): 353-&.

Schuster, S. C. (2008). "Next-generation sequencing transforms today's biology." Nature Methods **5**(1): 16-18.

Sequeira, T., D. Tavares, et al. (1996). "Evidence for circulating hemocyte proliferation in the shrimp *Penaeus japonicus*." Developmental and Comparative Immunology **20**(2): 97-104.

Shapiro, L., J. Love, et al. (2007). Adhesion molecules in the nervous system: Structural insights into function and diversity. Annual Review of Neuroscience. **30:** 451-474.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nature Biotechnology **26**(10): 1135-1145.

Shia, A. K. H., M. Glittenberg, et al. (2009). "Toll-dependent antimicrobial responses in Drosophila larval fat body require Spatzle secreted by haemocytes." Journal of Cell Science **122**(24): 4505-4515.

Simpson, E. H. (1949). "Measurement of Diversity." Nature **163**(4148): 688-688.

Skern, R., P. Frost, et al. (2005). "Relative transcript quantification by Quantitative PCR: Roughly right or precisely wrong?" Bmc Molecular Biology **6**.

Slusarczyk, M., P. Dawidowicz, et al. (2005). "Hide, rest or die: a light-mediated diapause response in Daphnia magna to the threat of fish predation." Freshwater Biology **50**(1): 141-146.

Smeds, L. and A. Kunstner (2011). "CONDETRI - A Content Dependent Read Trimmer for Illumina Data." PLoS One **6**(10).

Smith, P., J. Mwangi, et al. (2011). "Alternative splicing of the Anopheles gambiae Dscam gene in diverse Plasmodium falciparum infections." Malaria Journal **10**(1): 156.

Snounou, G., S. Viriyakosol, et al. (1993). "High-sensitivity of detection of human malaria parasites by the use of nested polymerase chain-reaction." Molecular and Biochemical Parasitology **61**(2): 315-320.

Soderhall, K. and L. Cerenius (1998). "Role of the prophenoloxidase-activating system in invertebrate immunity." Current Opinion in Immunology **10**(1): 23-28.

Srinivasan, K., L. Shiue, et al. (2005). "Detection and measurement of alternative splicing using splicing-sensitive microarrays." Methods **37**(4): 345-359.

Steiner, C. F. (2002). "Context-dependent effects of Daphnia pulex on pond ecosystem function: observational and experimental evidence." Oecologia **131**(4): 549-558.

Stirnadel, H. A. and D. Ebert (1997). "Prevalence, host specificity and impact on host fecundity of microparasites and epibionts in three sympatric Daphnia species." Journal of Animal Ecology **66**(2): 212-222.

Stroschein-Stevenson, S. L., E. Foley, et al. (2006). "Identification of Drosophila gene products required for phagocytosis of Candida albicans." Plos Biology **4**(1): 87-99.

Stuart, L. M. and R. A. Ezekowitz (2008). "Phagocytosis and comparative innate immunity: learning on the fly." Nature Reviews Immunology **8**(2): 131-141.

Su, X. Z. and T. E. Wellems (1996). "Toward a high-resolution *Plasmodium falciparum* linkage map: Polymorphic markers from hundreds of simple sequence repeats." Genomics **33**(3): 430-444.

Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0." Molecular Biology and Evolution **24**(8): 1596-1599.

Tamura, K., S. Subramanian, et al. (2004). "Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks." Molecular Biology and Evolution **21**(1): 36-44.

Tazi, J., N. Bakkour, et al. (2009). "Alternative splicing and disease." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease **1792**(1): 14-26.

Teixeira, L., A. Ferreira, et al. (2008). "The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in Drosophila melanogaster." Plos Biology **6**(12): 2753-2763.

Tinsley, M. C., S. Blanford, et al. (2006). "Genetic variation in Drosophila melanogaster pathogen susceptibility." Parasitology **132**: 767-773.

Traniello, J. F. A., R. B. Rosengaus, et al. (2002). "The development of immunity in a social insect: Evidence for the group facilitation of disease resistance." Proceedings of the National Academy of Sciences of the United States of America **99**(10): 6838-6842.

Tzou, P., E. De Gregorio, et al. (2002). "How Drosophila combats microbial infection: a model to study innate immunity and host-pathogen interactions." Current Opinion in Microbiology **5**(1): 102-110.

Ulvila, J., L.-M. Vanha-Aho, et al. (2011). "Drosophila phagocytosis - still many unknowns under the surface." Apmis **119**(10): 651-662.

Verspoor, R. L. and P. R. Haddrill (2011). "Genetic Diversity, Population Structure and Wolbachia Infection Status in a Worldwide Sample of Drosophila melanogaster and D. simulans Populations." PLoS One **6**(10).

Volkman, S. K., A. E. Barry, et al. (2001). "Recent origin of *Plasmodium falciparum* from a single progenitor." Science **293**(5529): 482-484.

Vulule, J. M., R. F. Beach, et al. (1994). "Reduced susceptibility of Anopheles gambiae to permethrin associated with the use of permethrin-impregnated bednets and curtains in Kenya." Medical and Veterinary Entomology **8**(1): 71-75.

Walliker, D., I. A. Quakyi, et al. (1987). "Genetic Analysis of the Human Malaria Parasite *Plasmodium falciparum*." Science **236**(4809): 1661-1666.

Watson, F. L., R. Puttmann-Holgado, et al. (2005). "Extensive diversity of Ig-superfamily proteins in the immune system of insects." Science **309**(5742): 1874-1878.

Watthanasurorot, A., P. Jiravanichpaisal, et al. (2011). "Bacteria-Induced Dscam Isoforms of the Crustacean, *Pacifastacus leniusculus*." PLoS Pathog **7**(6): e1002062.

Witteveldt, J., C. C. Cifuentes, et al. (2004). "Protection of *Penaeus monodon* against white spot syndrome virus by oral vaccination." Journal of Virology **78**(4): 2057-2061.

Wojtowicz, W. M., J. J. Flanagan, et al. (2004). "Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding." Cell **118**(5): 619-633.

Wong, Z. S., L. M. Hedges, et al. (2011). "Wolbachia-Mediated Antibacterial Protection and Immune Gene Regulation in Drosophila." PLoS One **6**(9).

Wooden, J., S. Kyes, et al. (1993). "PCR and Strain Identification in *Plasmodium falciparum*." Parasitology Today **9**(8): 303-305.

Yamakawa, K., Y. K. Huo, et al. (1998). "DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system." Human Molecular Genetics **7**(2): 227-237.

Zhang, Z. T. and S. Y. Zhu (2009). "Drosomycin, an essential component of antifungal defence in Drosophila." Insect Molecular Biology **18**(5): 549-556.
Zinn, K. (2007). "Dscam and neuronal uniqueness." Cell **129**(3): 455-456.

Appendices

7.1. Chapter 2 appendices

7.1.1. Neighbour-Joining tree construction and chi-square contingency table analysis

Methods: Full *AgDscam* exon variant DNA sequences were aligned using ClustalW and Neighbour-Joining tree constructed with MEGA (version 4) software (Tamura, Dudley et al. 2007).

Results:

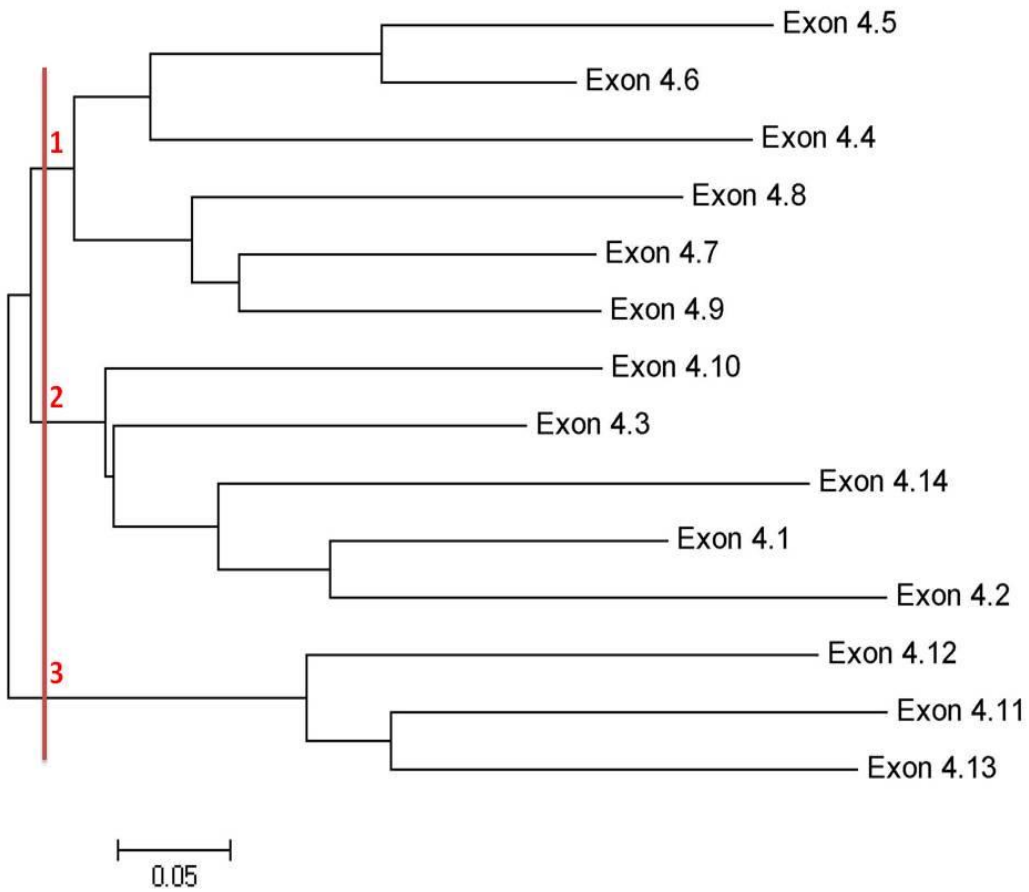


Figure S1: Neighbour-Joining tree of *AgDscam* exon 4 variants clustered into 3 groups. Full-length DNA sequences of the 14 variants of exon 4 were aligned using BioEdit software version 7.0.9.0, and the tree constructed using the MEGA (version 4) program.

Table S1a: **Cross tabulation and chi-square results for exon 4 clustered into 3 groups in our NJ tree between control and exposed treatments in the *field***. Analysis was carried out using Minitab® 15.1.1.0. Exon 4 variants 4.11, 4.12 and 4.13 are represented as Group 3 and were under-represented in control mosquitoes compared to exposed treatments in the field (Pearson Chi-Square = 6.318, DF = 2, p = 0.042).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	82	59	12
	Expected	79.7	52.3	21.0
Exposed	Observed	255	162	77
	Expected	257.3	168.7	68.0

Table S1b: **Cross tabulation and chi-square results for exon 4 clustered into 3 groups in our NJ tree between controls, single-, double- and triple-exposed treatments in the *field***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 8.970, DF = 6, p = 0.175).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	82	59	12
	Expected	79.7	52.3	21.1
Single	Observed	82	51	18
	Expected	78.7	51.6	20.8
Double	Observed	94	63	33
	Expected	99.0	64.9	26.1
Triple	Observed	79	48	26
	Expected	79.7	52.3	21.1

Table S1c: **Cross tabulation and chi-square results for exon 4 clustered into 3 groups in our NJ tree between control and exposed treatments in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-square = 3.231, DF = 2, p = 0.199).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	169	108	43
	Expected	159.3	121.6	39.0
Exposed	Observed	558	447	135
	Expected	567.7	433.4	139.0

Table S1d: **Cross tabulation and chi-square results for exon 4 clustered into 3 groups in our NJ tree between controls, 3D7, HB3 and mixed exposed treatments in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 4.724, DF = 6, p = 0.580).

Treatment		**Count**		
		Group 1	***Group 2***	***Group 3***
Control	Observed	169	108	43
	Expected	159.3	121.6	39.0
3D7	Observed	193	148	42
	Expected	190.7	145.6	46.7
HB3	Observed	173	144	49
	Expected	182.2	139.1	44.6
3D7/HB3	Observed	192	155	44
	Expected	194.7	148.6	47.7

Table S1e: **Cross tabulation and chi-square results for exon 4 clustered into 3 groups in our NJ tree between 3D7 and HB3 treatments in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 1.301, DF = 2, p = 0.522).

Treatment		**Count**		
		Group 1	***Group 2***	***Group 3***
3D7	Observed	193	148	42
	Expected	187.2	149.3	46.5
HB3	Observed	173	144	49
	Expected	178.8	142.7	44.5

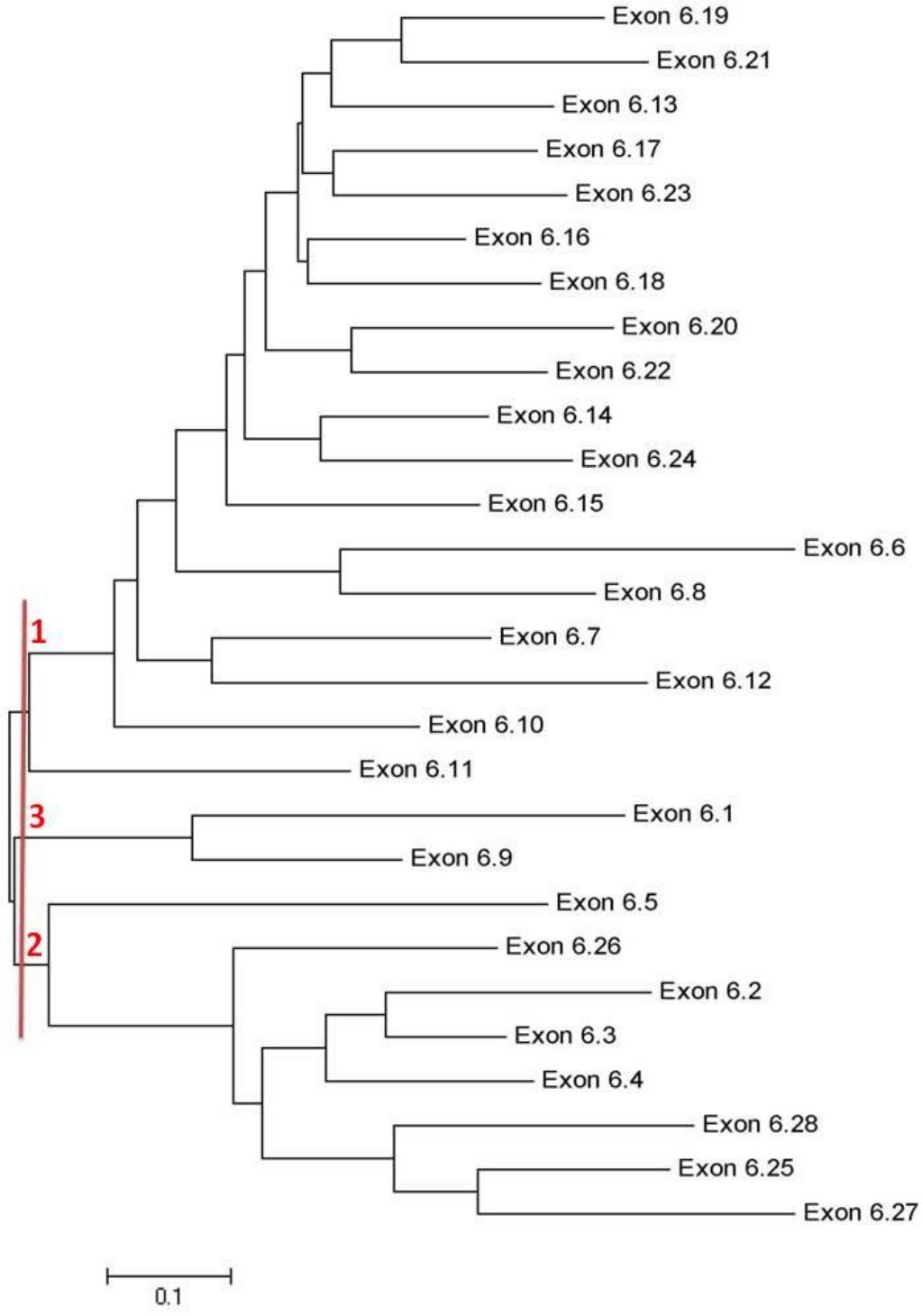


Figure S2: Neighbour-Joining tree of *AgDscam* exon 6 variants clustered into 3 groups. Full-length DNA sequences of 28 variants of exon 6 were aligned using BioEdit software version 7.0.9.0, and the tree constructed using the MEGA (version 4) program.

Table S2a: **Cross tabulation and chi-square results for exon 6 clustered into 3 groups in our NJ tree between control and exposed treatments in the *field***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 5.243, DF = 2, p = 0.073).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	105	29	19
	Expected	94.6	30.4	27.9
Exposed	Observed	315	106	105
	Expected	325.4	104.6	96.1

Table S2b: **Cross tabulation and chi-square results for exon 6 clustered into 3 groups in our NJ tree between controls, single-, double- and triple-exposed treatments in the *field***. Analysis was carried out using Minitab® 15.1.1.0. Exon 6 variants 6.1 and 6.9 are represented as Group 3 and appeared to be under-represented in control mosquitoes but over-represented in mosquitoes exposed to a single parasite genotype in the field (Pearson Chi-Square = 19.975, DF = 6, p = 0.003).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	105	29	19
	Expected	94.6	30.4	27.9
Single	Observed	87	25	47
	Expected	98.4	31.6	29.0
Double	Observed	122	43	31
	Expected	121.2	39.0	35.8
Triple	Observed	106	38	27
	Expected	105.8	34.0	31.2

Table S2c: **Cross tabulation and chi-square results for exon 6 clustered into 3 groups in our NJ tree between control and exposed treatments in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 1.446, DF = 2, p = 0.485).

Treatment		Count		
		Group 1	***Group 2***	***Group 3***
Control	Observed	182	56	68
	Expected	172.8	60.4	72.8
Exposed	Observed	599	217	261
	Expected	608.2	212.6	256.2

Table S2d: **Cross tabulation and chi-square results for exon 6 clustered into 3 groups in our NJ tree between controls, 3D7, HB3 and mixed exposures in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 3.984, DF = 6, p = 0.679).

Treatment		**Count**		
		Group 1	***Group 2***	***Group 3***
Control	Observed	182	56	68
	Expected	172.8	60.4	72.8
3D7	Observed	209	64	90
	Expected	205.0	71.6	86.4
HB3	Observed	189	71	79
	Expected	191.4	66.9	80.6
Mix	Observed	201	82	92
	Expected	211.8	74.0	89.2

Table S2e: **Cross tabulation and chi-square results for exon 6 clustered into 3 groups in our NJ tree between 3D7 and HB3 exposures in the *lab***. Analysis was carried out using Minitab® 15.1.1.0. Differences between observed and expected values were not statistically significant (Pearson Chi-Square = 1.265, DF = 2, p = 0.531).

Treatment		**Count**		
		Group 1	***Group 2***	***Group 3***
3D7	Observed	209	64	90
	Expected	205.8	69.8	87.4
HB3	Observed	189	71	79
	Expected	192.2	65.2	81.6

7.1.2. PCR detection of Plasmodium species

Methods: Existing primer sequences for *Plasmodium falciparum, P. malariae, P. ovale* and *P. vivax* 18s rRNA, and nested PCR conditions were taken from (Snounou, Viriyakosol et al. 1993). Template DNA was extracted from blood samples spotted onto Whatman® paper using a Chelex-100 isolation technique (Wooden, Kyes et al. 1993).

Results: All blood samples were infected with *P. falciparum*, no blood was infected with *P. ovale* or *P. vivax*, while two samples contained *P. malariae* in addition to *P. falciparum* (Table S3).

Table S3: Scoring of presence or absence of four species of *Plasmodium* following PCR detection.

Treatment	*P. falciparum*	*P. malariae*	*P.ovale*	*P.vivax*
SH158 (triple infection_#1)	+	-	-	-
SH93 (triple infection_#2)	+	-	-	-
SH79 (double infection_#1)	+	-	-	-
SH25 (double infection_#2)	+	-	-	-
K10(single infection_#1)	+	+	-	-
IG14 (single infection_#2)	+	+	-	-

Key: +; Detected, -; Not detected

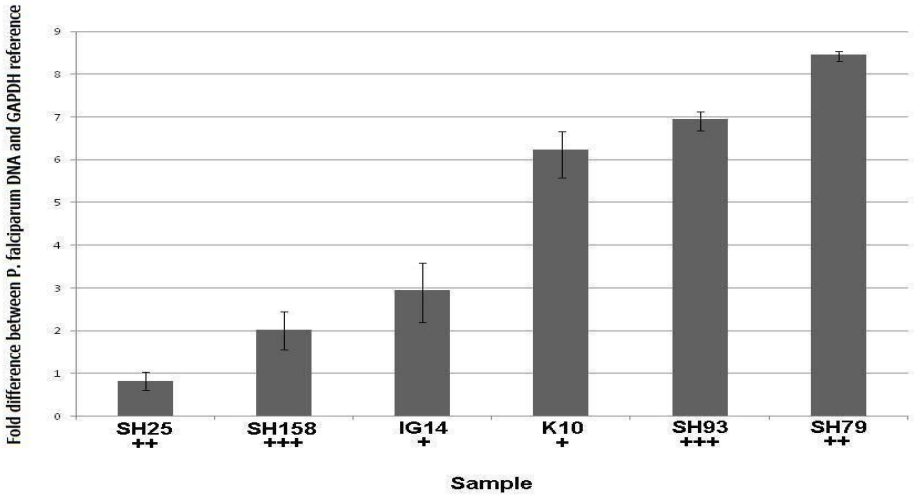
7.1.3. *Plasmodium falciparum* intensity

Methods: Real-time quantitative PCR assays were performed on an Applied Biosystems StepOnePlus cycler and we used the non-specific nucleic acid stain SYBR Green I which preferentially binds to dsDNA. For our reference gene, human-specific Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) primers were designed using Primer3 software (http://frodo.wi.mit.edu/) and checked for quality (tendency to dimerise, cross-dimerise, hairpin etc) with NetPrimer (http://www.premierbiosoft.com/netprimer/index.html) (primers were: F: 5' – CGACCACTTTGTCAAGCTCA – 3', R: 5' - GGTGGTCCAGGGGTCTTACT – 3') and amplified an 112bp fragment. Existing primers were used to amplify *Plasmodium falciparum* SSU rRNA (primers were: F: 5' – TCTAGGGGAACTATTTTAGCTT – 3', R: 5' – CACAGTAAATGCTTTAACTGTT – 3') isolating an 180bp fragment (Bell and Ranford-Cartwright 2004). All cycler runs included serial dilutions of parasite DNA to indicate primer efficiency and included a melting curve analysis for product identification. Each reaction contained: 8µl SYBR Green Master Mix, 10µl ddH_2O, 1µl primer mix (10mM concentration) and 1µl template genomic DNA. PCR conditions were: holding stage: 95ºC for 10 minutes; cycling stage: 40 cycles of 95ºC for 15 seconds, 55ºC for 60 seconds; and melt curve stage: 95ºC for 15 seconds, 55ºC for 60 seconds, temperature increment +0.3ºC, 95ºC for 15 seconds.

We quantified relative parasite intensity for two blood samples which were confirmed (using microsatellite loci sizing) to contain a minimum of one *P. falciparum* genotype (K10, IG14; single infections), two blood samples containing a minimum of two *P. falciparum* genotypes (SH79, SH25; double infections), and two blood samples containing a minimum of three *P. falciparum* genotypes (SH158, SH93; triple infections). Mean C_t values were calculated from 3 technical replicates per treatment. Delta C_t values were calculated (Delta C_t = Mean $C_{t,sample}$

– Mean $C_{t,reference}$), and we plotted fold difference between target DNA and reference DNA ($2^{-delta\ Ct}$) (Figure S3).

Results:



Key: +++; triple infection, ++; double infection, +; single infection.

Figure S3: Fold difference in relative DNA quantity between human reference gene (GAPDH) and target (*Plasmodium falciparum*). Error bars represent standard error and calculated from three independent replicates.

7.2. Chapter 3 appendices

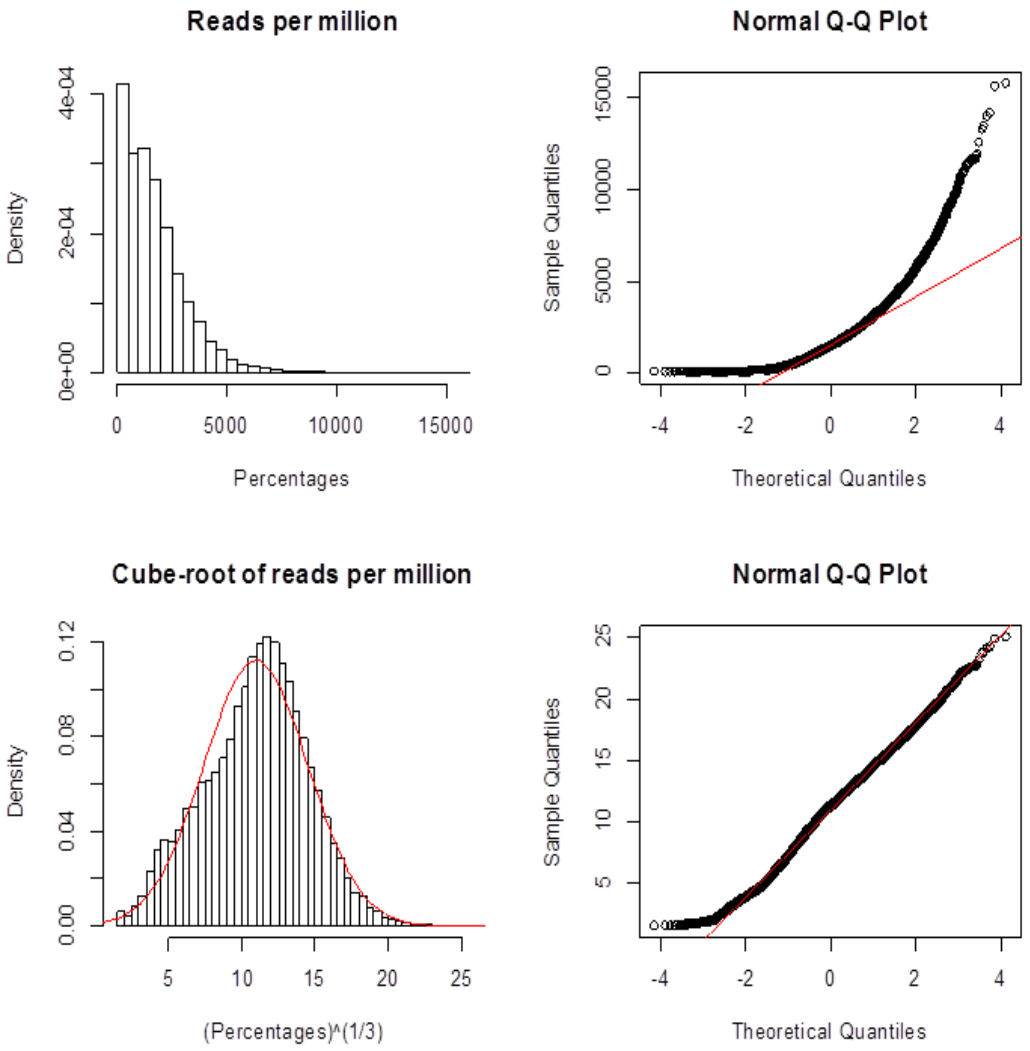


Figure S4: Quantile-Quantile (Q-Q) plot of Illumina sequencing raw count data of Chapter 3 before transformation (top) and after cube-root of percentage transformation (bottom).

Table S4: Simpson's Index of Diversity values for all treatments in the study (Chapter 3).

Diversity indices for exon 4, exon 6, and exon 4-6 combinations are shown and were calculated as Simpson's Index of Diversity (D) = 1-sum((Exon*i*/sum(Exon*i*))^2).

Ref	Treatment	Exon 4 (1-D)	Exon 6 (1-D)	Exon 4-6 (1-D)
R01	UNW_wol+_6H_3	0.8993	0.9723	0.9970
R02	UNW_wol+_30H_3	0.8939	0.9730	0.9967
R03	RING_wol+_6H_3	0.9019	0.9716	0.9967
R04	RING_wol+_30H_3	0.9061	0.9724	0.9970
R05	DCV_wol+_6H_3	0.8986	0.9723	0.9967
R06	DCV_wol+_30H_3	0.8992	0.9737	0.9973
R07	UNW_wol-_6H_3	0.8982	0.9731	0.9971
R08	UNW_wol-_30H_3	0.9023	0.9738	0.9972
R09	RING_wol-_6H_3	0.8933	0.9734	0.9971
R10	RING_wol-_30H_3	0.8983	0.9722	0.9967
R11	OIL_wol-_6H_3	0.9021	0.9716	0.9970
R12	OIL_wol-_30H_3	0.8999	0.9721	0.9967
R13	BASS_wol-_6H_3	0.8997	0.9724	0.9970
R14	BASS_wol-_30H_3	0.9010	0.9715	0.9970
R15	DCV_wol-_6H_3	0.9041	0.9716	0.9970
R16	DCV_wol-_30H_3	0.9023	0.9733	0.9973
R17	PiGV_wol-_6H_3	0.8964	0.9728	0.9967
R18	PiGV_wol-_30H_3	0.9020	0.9728	0.9970
R19	Ecoli_wol-_6H_3	0.8991	0.9722	0.9969
R20	Ecoli_wol-_30H_3	0.9051	0.9715	0.9971
R21	Mlut_wol-_6H_3	0.9017	0.9725	0.9972
R22	Mlut_wol-_30H_3	0.9023	0.9732	0.9972
R23	UNW_wol+_6H_4	0.8966	0.9717	0.9970
R24	UNW_wol+_30H_4	0.8930	0.9732	0.9969
R25	RING_wol+_6H_4	0.9016	0.9721	0.9971
R26	RING_wol+_30H_4	0.8983	0.9729	0.9971
R27	DCV_wol+_6H_4	0.9028	0.9724	0.9971
R28	DCV_wol+_30H_4	0.9031	0.9727	0.9971
R29	UNW_wol-_6H_4	0.8929	0.9721	0.9966
R30	UNW_wol-_30H_4	0.8966	0.9728	0.9969
R31	RING_wol-_6H_4	0.8991	0.9726	0.9971
R32	RING_wol-_30H_4	0.8975	0.9724	0.9970
R33	OIL_wol-_6H_4	0.8977	0.9733	0.9971
R34	OIL_wol-_30H_4	0.9006	0.9732	0.9973
R35	BASS_wol-_6H_4	0.9040	0.9726	0.9973
R36	BASS_wol-_30H_4	0.9053	0.9725	0.9972
R37	DCV_wol-_6H_4	0.8956	0.9733	0.9972
R38	DCV_wol-_30H_4	0.9003	0.9733	0.9972
R39	PiGV_wol-_6H_4	0.8872	0.9724	0.9963
R40	PiGV_wol-_30H_4	0.8990	0.9718	0.9967
R41	Ecoli_wol-_6H_4	0.8982	0.9733	0.9972
R42	Ecoli_wol-_30H_4	0.8968	0.9724	0.9969
R43	Mlut_wol-_6H_4	0.8977	0.9725	0.9970
R44	Mlut_wol-_30H_4	0.9010	0.9729	0.9971

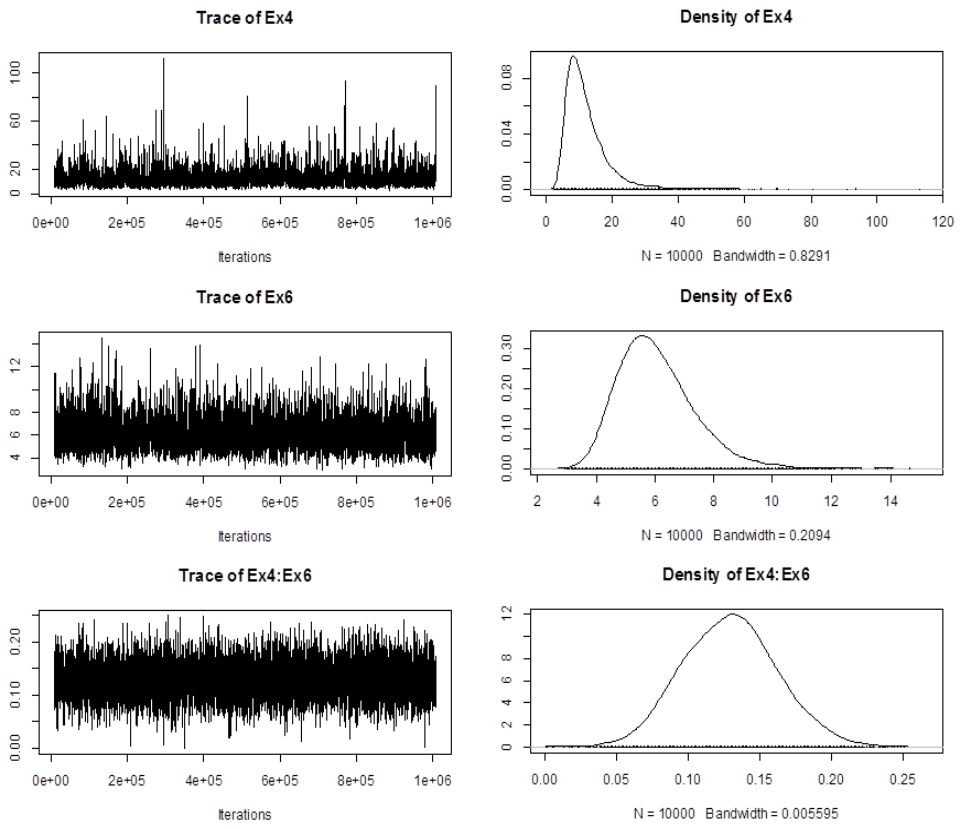


Figure S5: Mixing and posterior distributions of significant effects found in model testing for effects of wounding and time on *D. melanogaster* Dscam expression. Results are outlined in Table 1 of Chapter 3. Mixing was generally good (left panels), and unimodal estimates were clear (right panels).

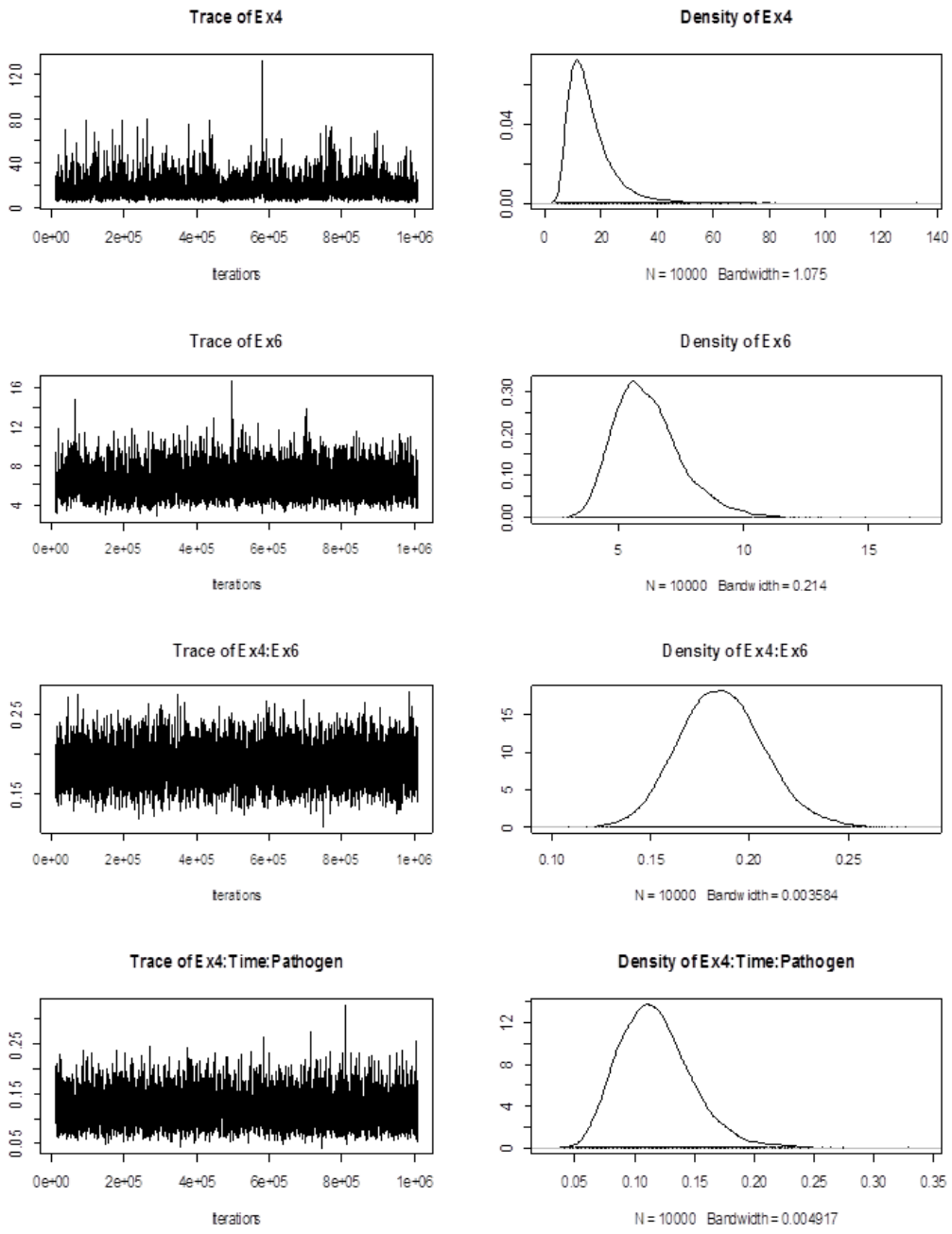


Figure S6: Mixing and posterior distributions of significant effects found in model testing for effects of pathogen-exposure and time on *D. melanogaster* Dscam expression. Results are outlined in Table 2 of Chapter 3.

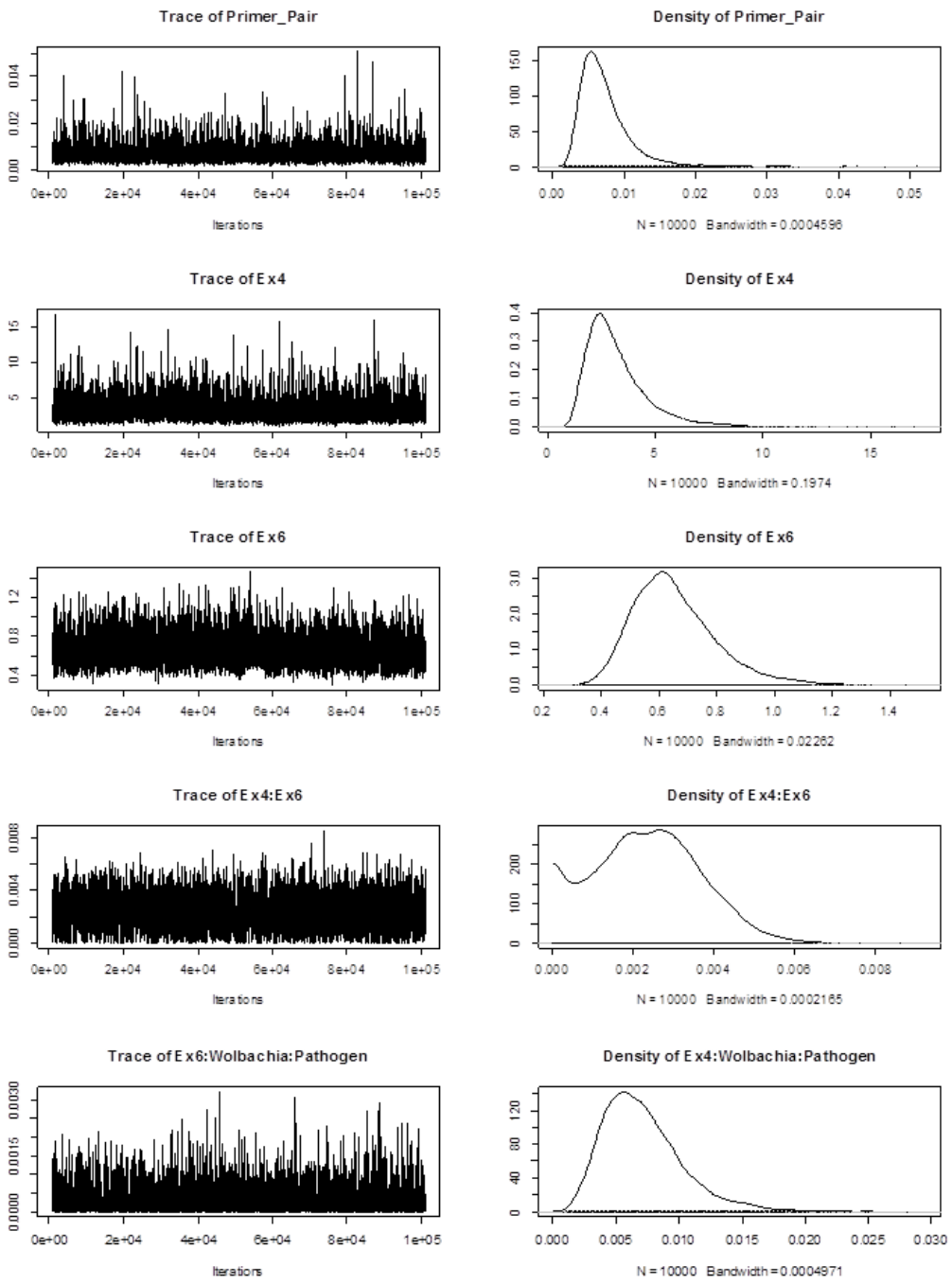


Figure S7: Mixing and posterior distributions of significant effects found in model testing for effects of DCV-exposure on *D. melanogaster* Dscam expression in *Wolbachia*-cleared and *Wolbachia*-infected treatments. Results are outlined in Table 3 of Chapter 3.

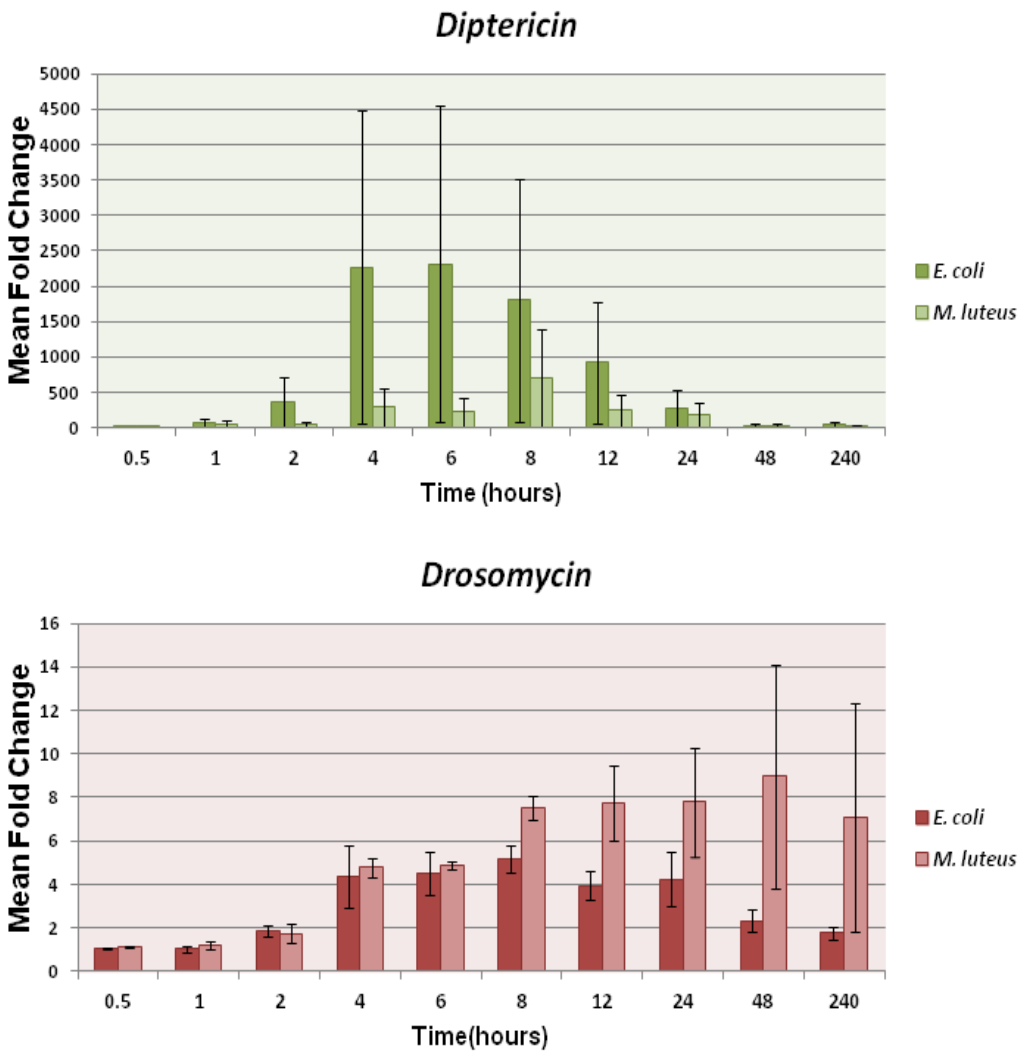


Figure S8: Mean fold change of Diptericin and Drosomycin expression in bacteria-exposed flies relative to controls. Diptericin levels peaked in response to both *E. coli* and *M. luteus* between 4-8 hours post-exposure (top). Drosomycin levels steadily increased following exposure to *M. luteus* until a peak of around 48 hours, while levels peaked around 8 hours following exposure to *E. coli* (bottom). Error bars represent SEM (n=3).

7.3. Chapter 4 appendices

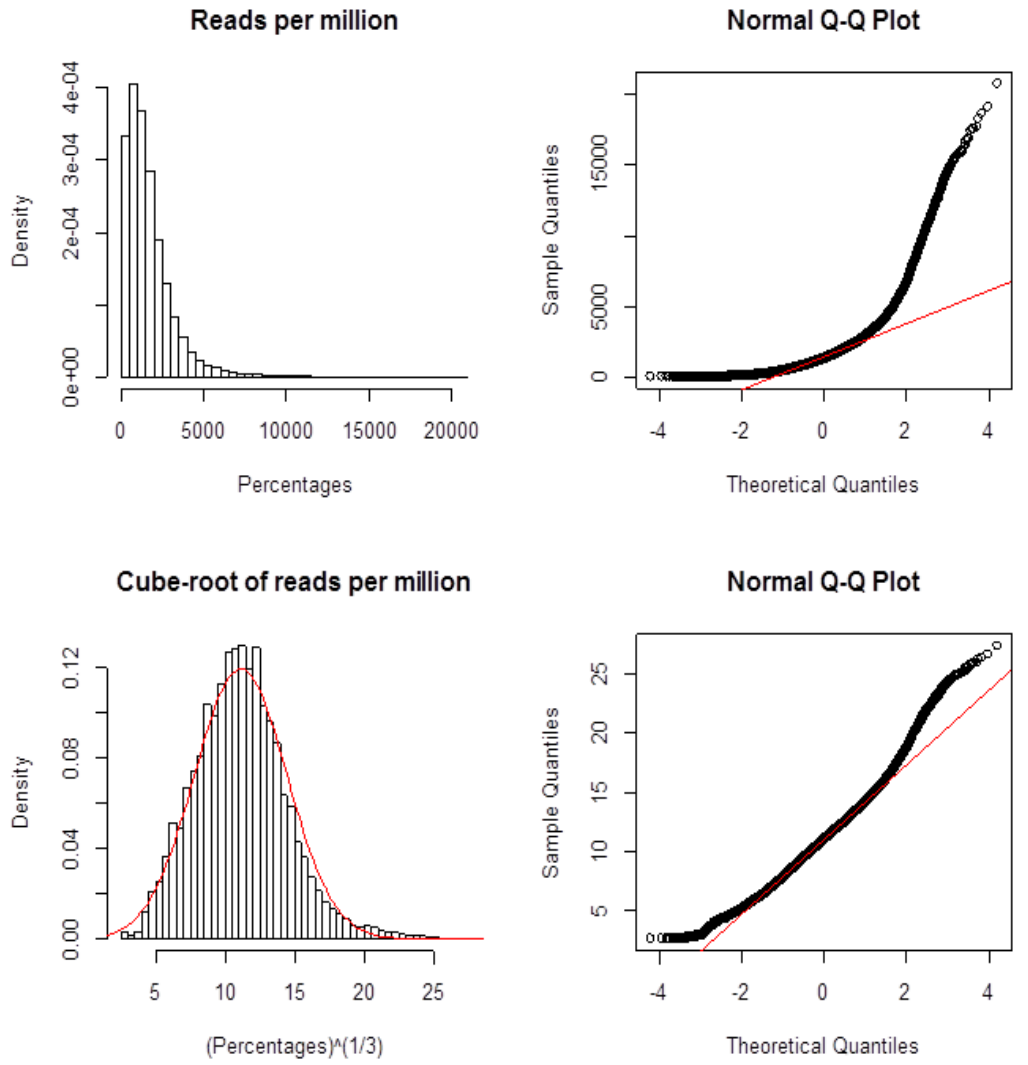


Figure S9: Quantile-Quantile (Q-Q) plot of Illumina sequencing raw count data of Chapter 4 before transformation (top) and after cube-root of percentage transformation (bottom).

Table S5: Significance of departure from even expression of exon 4-6 combinations in each tissue of unwounded *D. melanogaster* (Chapter 4). P-values are from Fisher's Exact Tests and significant p-values are asterisked.

Ref	Treatment	p-value
1	UNW_1_Head	0.274
2	UNW_1_Body	0.021*
3	UNW_1_Cells	0.326
4	RING_1_Head	0.149
5	RING_1_Body	0.299
6	RING_1_Cells	0.525
7	Ecoli_1_Head	0.692
8	Ecoli_1_Body	0.369
9	Ecoli_1_Cells	0.409
10	Mlut_1_Head	0.004*
11	Mlut_1_Body	0.347
12	Mlut_1_Cells	0.459
13	UNW_2_Head	0.353
14	UNW_2_Body	0.193
15	UNW_2_Cells	0.386
16	RING_2_Head	0.334
17	RING_2_Body	0.059
18	RING_2_Cells	0.245
19	Ecoli_2_Head	0.061
20	Ecoli_2_Body	0.078
21	Ecoli_2_Cells	0.037*
22	Mlut_2_Head	0.498
23	Mlut_2_Body	0.441
24	Mlut_2_Cells	0.242
25	UNW_3_Head	0.704
26	UNW_3_Body	0.806
27	UNW_3_Cells	0.204
28	RING_3_Head	0.003*
29	RING_3_Body	0.519
30	RING_3_Cells	0.088
31	Ecoli_3_Head	0.022*
32	Ecoli_3_Body	0.239
33	Ecoli_3_Cells	0.570
34	Mlut_3_Head	0.0006*
35	Mlut_3_Body	0.467
36	Mlut_3_Cells	0.856
37	UNW_4_Head	0.185
38	UNW_4_Body	0.218
39	UNW_4_Cells	0.070
40	RING_4_Head	0.634
41	RING_4_Body	0.372

42	RING_4_Cells	0.344
43	Ecoli_4_Head	0.535
44	Ecoli_4_Body	0.316
45	Ecoli_4_Cells	0.032*
46	Mlut_4_Head	0.255
47	Mlut_4_Body	0.848
48	Mlut_4_Cells	0.163
49	UNW_5_Head	0.020*
50	UNW_5_Body	0.909
51	UNW_5_Cells	0.675
52	RING_5_Head	0.143
53	RING_5_Body	0.877
54	RING_5_Cells	0.065
55	Ecoli_5_Head	0.446
56	Ecoli_5_Body	0.299
57	Ecoli_5_Cells	0.477
58	Mlut_5_Head	0.027*
59	Mlut_5_Body	0.075
60	Mlut_5_Cells	0.244
61	UNW_6_Head	0.881
62	UNW_6_Body	0.169
63	UNW_6_Cells	0.503
64	RING_6_Head	0.205
65	RING_6_Body	0.070
66	RING_6_Cells	0.516
67	Ecoli_6_Head	0.532
68	Ecoli_6_Body	0.452
69	Ecoli_6_Cells	0.514
70	Mlut_6_Head	0.473
71	Mlut_6_Body	0.201
72	Mlut_6_Cells	0.843

Table S6: Simpson's Index of Diversity values for all treatments in the study (Chapter 4).

Diversity indices for exon 4, exon 6, and exon 4-6 combinations are shown and were calculated as Simpson's Index of Diversity (D) = 1-sum((Exon*i*/sum(Exon*i*))^2).

Ref	Treatment	Exon 4 (1-D)	Exon 6 (1-D)	Exon 4-6 (1-D)
R01	UNW_1_Head	0.9042	0.9657	0.9967
R02	UNW_1_Body	0.8950	0.9645	0.9961
R03	UNW_1_Cells	0.9034	0.9663	0.9967
R04	RING_1_Head	0.9033	0.9656	0.9967
R05	RING_1_Body	0.8954	0.9656	0.9964
R06	RING_1_Cells	0.9028	0.9660	0.9966
R07	Ecoli_1_Head	0.9030	0.9658	0.9967
R08	Ecoli_1_Body	0.8966	0.9662	0.9964
R09	Ecoli_1_Cells	0.9040	0.9656	0.9966
R10	Mlut_1_Head	0.9032	0.9659	0.9967
R11	Mlut_1_Body	0.8964	0.9655	0.9964
R12	Mlut_1_Cells	0.9026	0.9637	0.9963
R13	UNW_2_Head	0.9046	0.9659	0.9967
R14	UNW_2_Body	0.8978	0.9659	0.9964
R15	UNW_2_Cells	0.9004	0.9661	0.9966
R16	RING_2_Head	0.9042	0.9659	0.9967
R17	RING_2_Body	0.8947	0.9641	0.9958
R18	RING_2_Cells	0.9061	0.9659	0.9968
R19	Ecoli_2_Head	0.9045	0.9660	0.9967
R20	Ecoli_2_Body	0.8971	0.9655	0.9964
R21	Ecoli_2_Cells	0.8989	0.9659	0.9963
R22	Mlut_2_Head	0.9044	0.9656	0.9967
R23	Mlut_2_Body	0.8916	0.9636	0.9955
R24	Mlut_2_Cells	0.9021	0.9660	0.9966
R25	UNW_3_Head	0.9053	0.9658	0.9967
R26	UNW_3_Body	0.8963	0.9650	0.9963
R27	UNW_3_Cells	0.9057	0.9664	0.9968
R28	RING_3_Head	0.9051	0.9652	0.9967
R29	RING_3_Body	0.8963	0.9647	0.9962
R30	RING_3_Cells	0.9046	0.9660	0.9967
R31	Ecoli_3_Head	0.9033	0.9661	0.9967
R32	Ecoli_3_Body	0.8923	0.9661	0.9963
R33	Ecoli_3_Cells	0.9062	0.9666	0.9968
R34	Mlut_3_Head	0.9044	0.9656	0.9967
R35	Mlut_3_Body	0.8975	0.9656	0.9964
R36	Mlut_3_Cells	0.9061	0.9660	0.9967
R37	UNW_4_Head	0.9054	0.9658	0.9968
R38	UNW_4_Body	0.8968	0.9656	0.9964
R39	UNW_4_Cells	0.9025	0.9647	0.9964
R40	RING_4_Head	0.9029	0.9661	0.9967
R41	RING_4_Body	0.8973	0.9655	0.9960

R42	RING_4_Cells	0.9045	0.9661	0.9967
R43	Ecoli_4_Head	0.9007	0.9653	0.9965
R44	Ecoli_4_Body	0.8958	0.9658	0.9962
R45	Ecoli_4_Cells	0.9018	0.9659	0.9966
R46	Mlut_4_Head	0.9028	0.9657	0.9966
R47	Mlut_4_Body	0.8976	0.9650	0.9960
R48	Mlut_4_Cells	0.9057	0.9662	0.9966
R49	UNW_5_Head	0.9059	0.9654	0.9967
R50	UNW_5_Body	0.9012	0.9657	0.9965
R51	UNW_5_Cells	0.9059	0.9656	0.9967
R52	RING_5_Head	0.9062	0.9656	0.9968
R53	RING_5_Body	0.8971	0.9654	0.9964
R54	RING_5_Cells	0.9048	0.9658	0.9967
R55	Ecoli_5_Head	0.9058	0.9657	0.9967
R56	Ecoli_5_Body	0.8956	0.9653	0.9963
R57	Ecoli_5_Cells	0.9044	0.9665	0.9967
R58	Mlut_5_Head	0.8961	0.9659	0.9964
R59	Mlut_5_Body	0.8977	0.9654	0.9964
R60	Mlut_5_Cells	0.9030	0.9655	0.9966
R61	UNW_6_Head	0.9049	0.9654	0.9967
R62	UNW_6_Body	0.8958	0.9662	0.9964
R63	UNW_6_Cells	0.9026	0.9662	0.9966
R64	RING_6_Head	0.9040	0.9667	0.9967
R65	RING_6_Body	0.8979	0.9649	0.9963
R66	RING_6_Cells	0.9050	0.9652	0.9965
R67	Ecoli_6_Head	0.9059	0.9656	0.9967
R68	Ecoli_6_Body	0.8953	0.9664	0.9964
R69	Ecoli_6_Cells	0.9004	0.9655	0.9965
R70	Mlut_6_Head	0.9052	0.9658	0.9967
R71	Mlut_6_Body	0.8972	0.9667	0.9965
R72	Mlut_6_Cells	0.9042	0.9669	0.9968

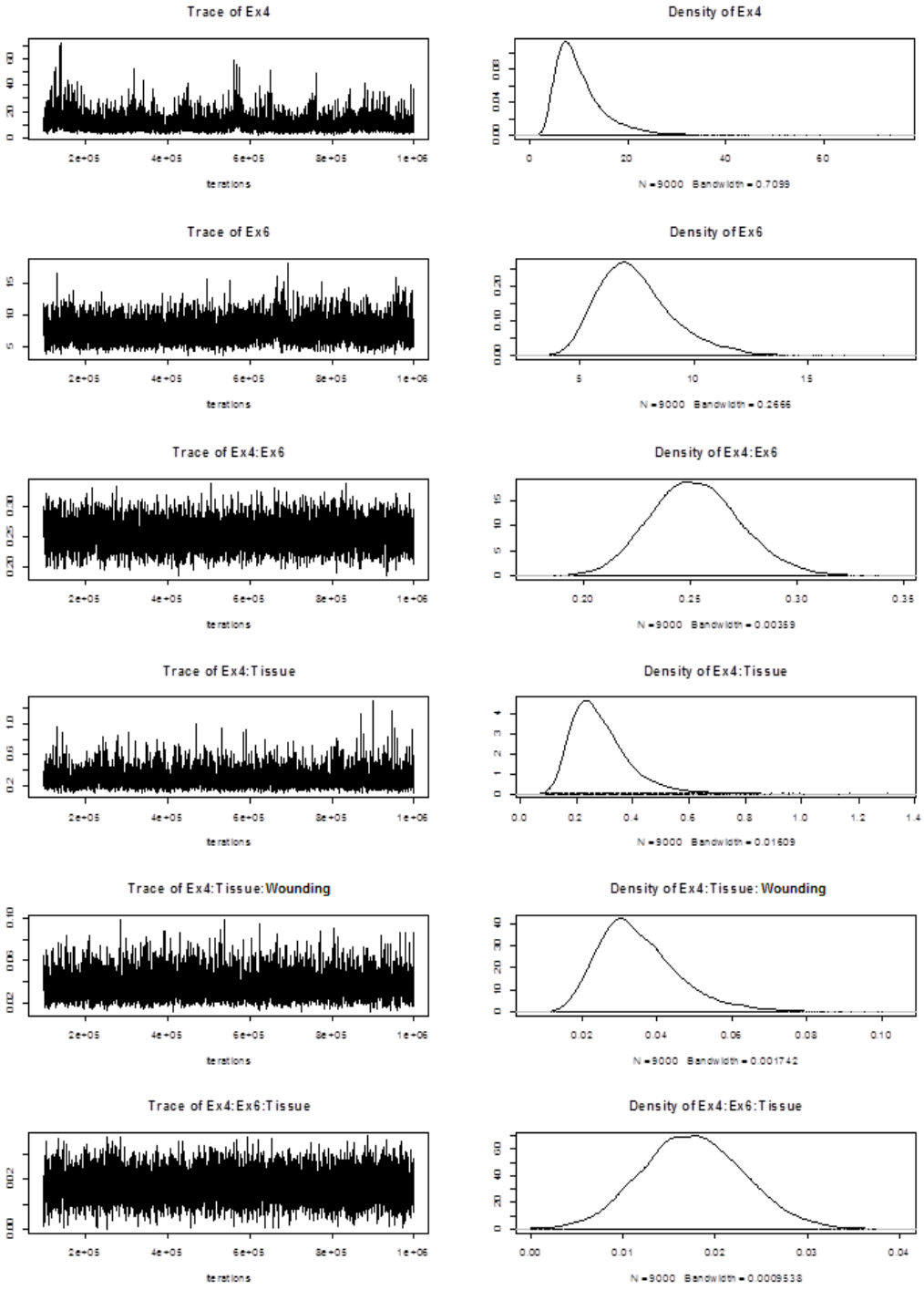


Figure S10: Mixing and posterior distributions of significant effects found in model testing for effects of wounding on *D. melanogaster* Dscam expression. Results are outlined in Table 1 of Chapter 4.

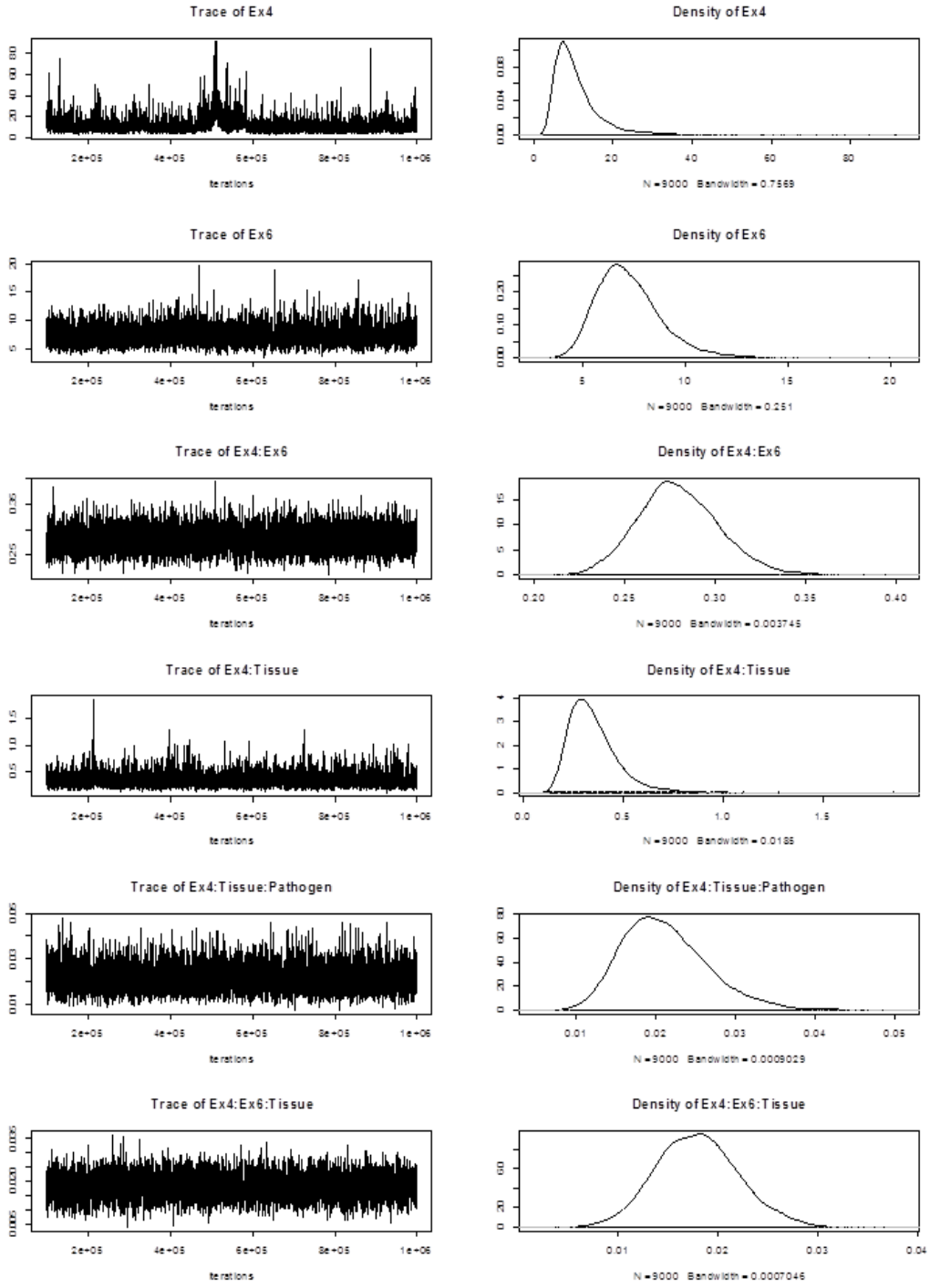


Figure S11: Mixing and posterior distributions of significant effects found in model testing for effects of pathogen-exposure on *D. melanogaster* Dscam expression. Results are outlined in Table 2 of Chapter 4.

7.4. Chapter 5 appendices

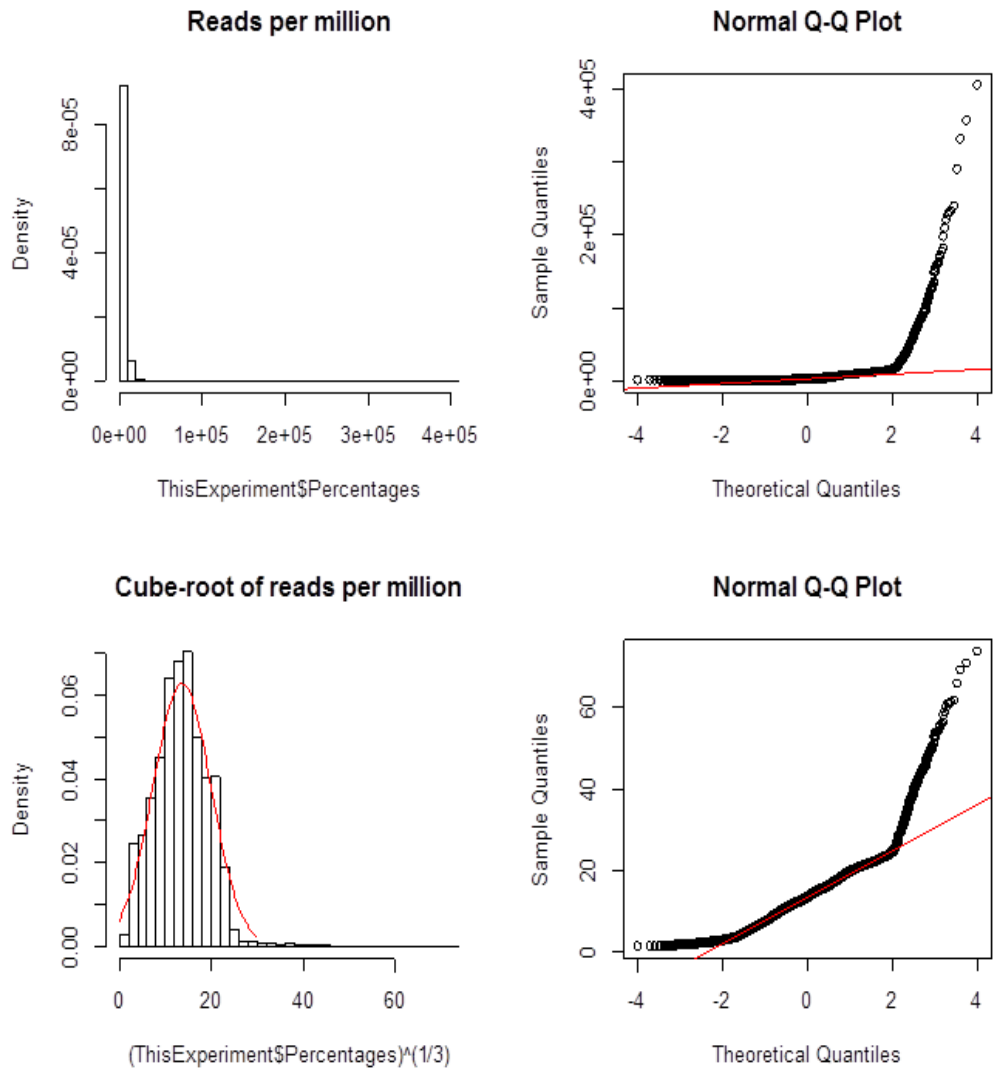


Figure S12: Quantile-Quantile (Q-Q) plot of Illumina sequencing raw count data of Chapter 5 before transformation (top) and after cube-root of percentage transformation (bottom).

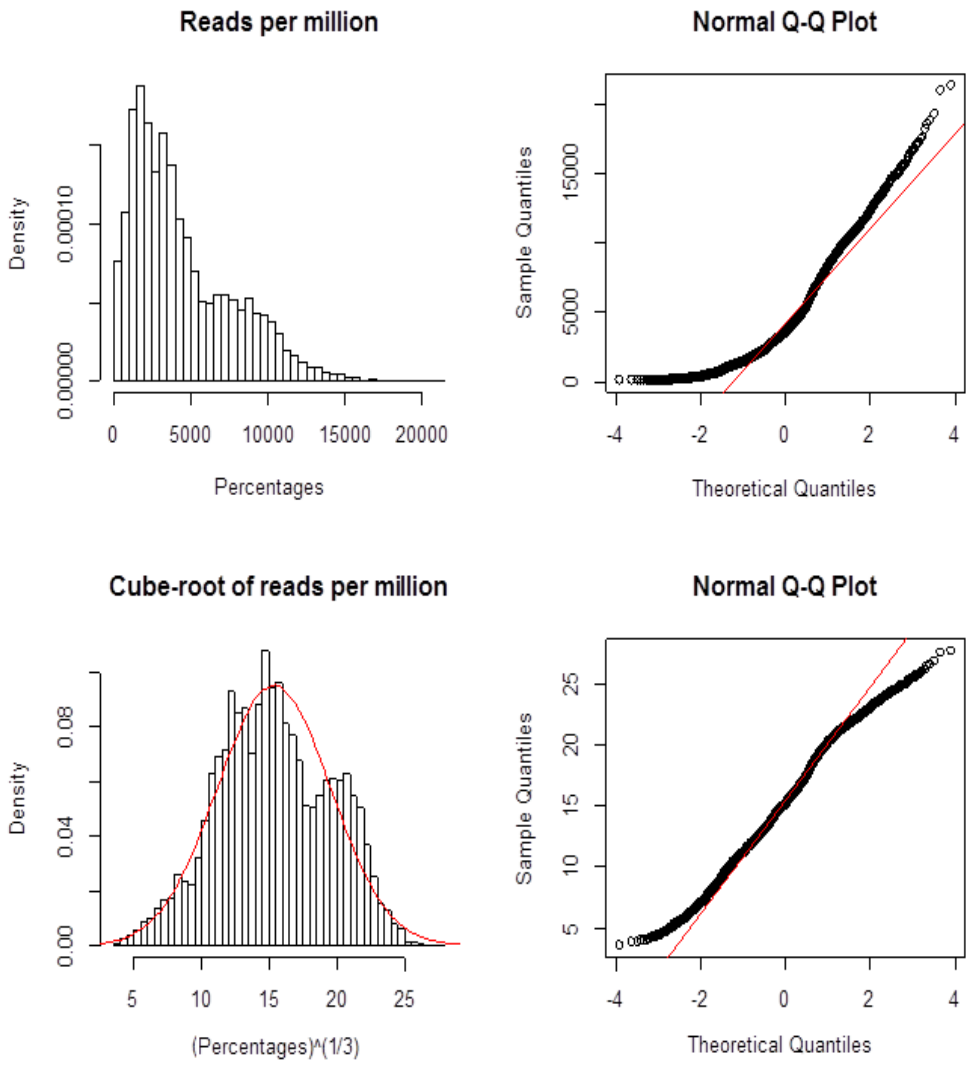


Figure S13: Quantile-Quantile (Q-Q) plot of Illumina sequencing raw count data of Chapter 5 *excluding haemolymph treatments* before transformation (top) and after cube-root of percentage transformation (bottom).

Table S7: Simpson's Index of Diversity values for all treatments in the study (Chapter 5).

Diversity indices for exon 4, exon 6, and exon 4-6 combinations are shown and were calculated as Simpson's Index of Diversity (D) = 1-sum((Exon*i*/sum(Exon*i*))^2).

Ref	Treatment	Exon 4 (1-D)	Exon 6 (1-D)	Exon 4-6 (1-D)
R41	Control_Carc_13	0.8530	0.9558	0.9934
R47	Control_Carc_14	0.8504	0.9559	0.9933
R59	Control_Carc_15	0.8538	0.9553	0.9934
R69	Control_Carc_16	0.8500	0.9556	0.9932
R03	Control_Carc_3	0.8541	0.9561	0.9935
R12	Control_Carc_5	0.8594	0.9558	0.9937
R23	Control_Carc_8	0.8563	0.9551	0.9935
R30	Control_Carc_9	0.8597	0.9553	0.9937
R38	Control_Cells_13	0.6819	0.7641	0.8856
R52	Control_Cells_14	0.7707	0.8921	0.9501
R56	Control_Cells_15	0.8393	0.9013	0.9552
R70	Control_Cells_16	0.4631	0.7629	0.7689
R06	Control_Cells_3	0.8383	0.9188	0.9559
R14	Control_Cells_5	0.7715	0.9027	0.9438
R26	Control_Cells_8	0.8231	0.9074	0.9434
R34	Control_Cells_9	0.8330	0.9300	0.9825
R45	Control_Gut_13	0.8459	0.9552	0.9930
R49	Control_Gut_14	0.8492	0.9545	0.9930
R62	Control_Gut_15	0.8459	0.9546	0.9928
R65	Control_Gut_16	0.8421	0.9551	0.9928
R02	Control_Gut_3	0.8431	0.9549	0.9928
R13	Control_Gut_5	0.8449	0.9550	0.9929
R21	Control_Gut_8	0.8443	0.9549	0.9928
R29	Control_Gut_9	0.8440	0.9548	0.9928
R39	Resistant_Carc_13	0.8524	0.9558	0.9934
R50	Resistant_Carc_14	0.8511	0.9555	0.9933
R60	Resistant_Carc_15	0.8530	0.9556	0.9934
R71	Resistant_Carc_16	0.8493	0.9556	0.9932
R01	Resistant_Carc_3	0.8565	0.9558	0.9936
R11	Resistant_Carc_5	0.8587	0.9557	0.9937
R19	Resistant_Carc_8	0.8562	0.9551	0.9934
R31	Resistant_Carc_9	0.8517	0.9555	0.9933
R44	Resistant_Cells_13	0.7530	0.8293	0.8638
R53	Resistant_Cells_14	0.6272	0.8299	0.8479
R57	Resistant_Cells_15	0.8240	0.9305	0.9752
R66	Resistant_Cells_16	0.7696	0.8704	0.9167
R05	Resistant_Cells_3	0.7906	0.9030	0.9322
R18	Resistant_Cells_5	0.6502	0.7867	0.7990
R27	Resistant_Cells_8	0.7922	0.8530	0.9229
R32	Resistant_Cells_9	0.8315	0.9406	0.9733
R40	Resistant_Gut_13	0.8401	0.9550	0.9925

R54	Resistant_Gut_14	0.8463	0.9542	0.9928
R61	Resistant_Gut_15	0.8477	0.9547	0.9930
R72	Resistant_Gut_16	0.8463	0.9551	0.9929
R04	Resistant_Gut_3	0.8454	0.9547	0.9929
R16	Resistant_Gut_5	0.8460	0.9547	0.9929
R20	Resistant_Gut_8	0.8501	0.9551	0.9931
R28	Resistant_Gut_9	0.8448	0.9542	0.9928
R43	Susceptible_Carc_13	0.8532	0.9558	0.9934
R51	Susceptible_Carc_14	0.8554	0.9556	0.9935
R58	Susceptible_Carc_15	0.8511	0.9549	0.9932
R67	Susceptible_Carc_16	0.8503	0.9557	0.9933
R07	Susceptible_Carc_3	0.8536	0.9559	0.9935
R10	Susceptible_Carc_5	0.8539	0.9559	0.9935
R24	Susceptible_Carc_8	0.8562	0.9554	0.9935
R33	Susceptible_Carc_9	0.8566	0.9555	0.9935
R37	Susceptible_Cells_13	0.5870	0.8453	0.8535
R48	Susceptible_Cells_14	0.8512	0.9481	0.9909
R64	Susceptible_Cells_16	0.7685	0.9180	0.9437
R08	Susceptible_Cells_3	0.8215	0.8676	0.9064
R17	Susceptible_Cells_5	0.6826	0.8357	0.8517
R25	Susceptible_Cells_8	0.7758	0.8847	0.8948
R35	Susceptible_Cells_9	0.8120	0.9072	0.9325
R42	Susceptible_Gut_13	0.8434	0.9550	0.9928
R46	Susceptible_Gut_14	0.8448	0.9545	0.9927
R63	Susceptible_Gut_15	0.8438	0.9549	0.9928
R68	Susceptible_Gut_16	0.8461	0.9548	0.9929
R09	Susceptible_Gut_3	0.8440	0.9549	0.9928
R15	Susceptible_Gut_5	0.8455	0.9547	0.9928
R22	Susceptible_Gut_8	0.8455	0.9546	0.9928
R36	Susceptible_Gut_9	0.8431	0.9546	0.9927

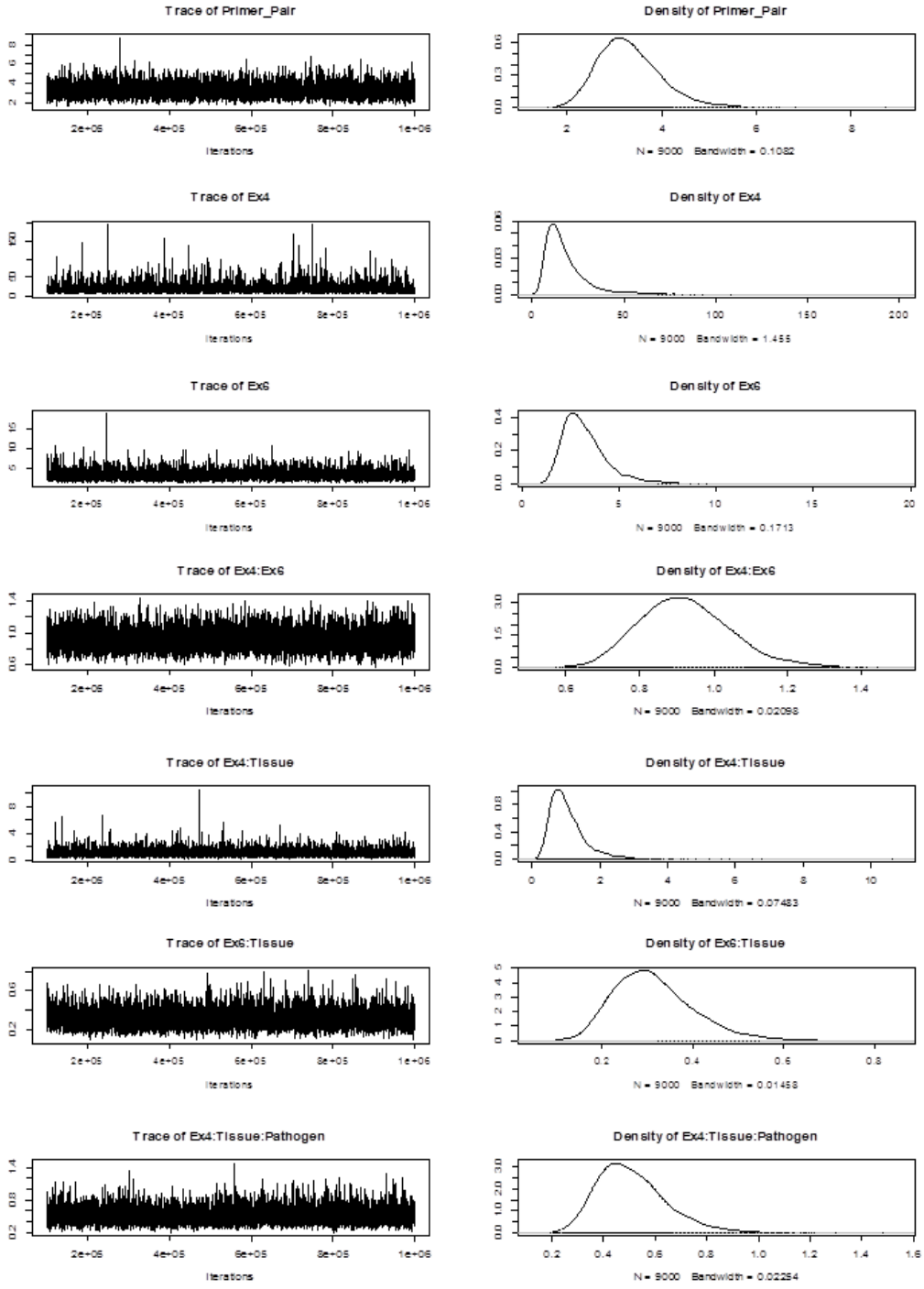


Figure S14: Mixing and posterior distributions of significant effects found in model testing for effects of parasite exposure on *D. magna* Dscam expression. Results are outlined in Table 1 of Chapter 5.

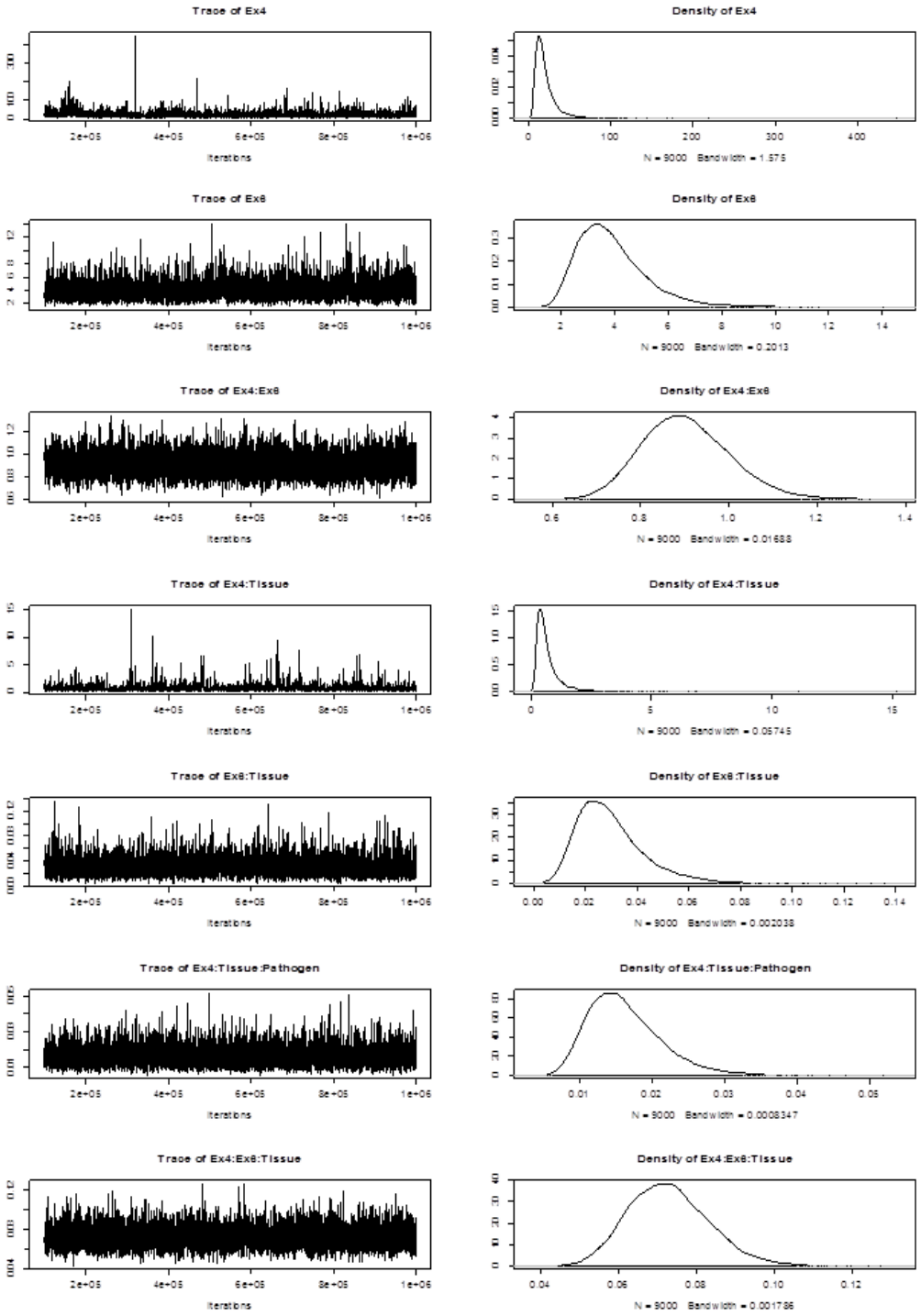


Figure S15: Mixing and posterior distributions of significant effects found in model testing for effects of pathogen-exposure on *D. magna* Dscam expression. Haemolymph treatments are excluded. Results are outlined in Table 2 of Chapter 5.